

## SYSTEM AND METHODS FOR NUCLEIC ACID AND POLYPEPTIDE SELECTION

### Field of the Invention

The present invention relates generally to compositions and methods for the identification and selection of nucleic acids and polypeptides.

### Background of the Invention

Ligand-receptor interactions are of interest for many reasons, from elucidating basic biological site recognition mechanisms to drug screening and rational drug design. It has been possible for many years to drive *in vitro* evolution of nucleic acids by selecting molecules out of large populations that preferentially bind to a selected target, then amplifying and mutating them for subsequent re-selection (Tuerk and Gold, Science 249:505 (1990), herein incorporated by reference).

The ability to perform such a selection process with proteins would be extremely useful. This would permit *in vitro* design and production of proteins that bind specifically to chosen ligands. The use of proteins, as compared to nucleic acids, is particularly advantageous because the twenty diverse amino acid side chains in proteins have far more binding possibilities than the four similar chains in nucleic acid side. Further, many biologically and medically relevant ligands bind proteins.

Both nucleic acid and protein evolution methods require access to a large and highly varied population of test molecules, a way to select members of the population that exhibit the desired properties, and the ability to reproduce the selected molecules with mutated variations to obtain another large population for subsequent selection.

Thus, a need exists for an *in vitro* nucleic acid-based protein evolution system that does not necessarily require initial knowledge of the nucleic acid's sequence or repeated chemical modification of the nucleic acids, and which can accurately link a mRNA to its protein.

### Summary of the Invention

Embodiments of present invention provides compositions and methods to select and evolve desired properties of proteins and nucleic acids. In various embodiments, the current invention provides modified tRNA's and tRNA analogs. Other embodiments include methods for generating polypeptides, assays enabling selection of

individual members of the population of polypeptides having desired characteristics, methods for amplifying the nucleic acids encoding such selected polypeptides, and methods for generating new variants to screen for enhanced properties.

In several embodiments, the present invention permits the attachment of a protein to its message without requiring modification of native mRNA, although modified mRNA may still be used. The specificity of the methods embodied in various aspects of the current invention are determined by the specificity of the codon-anticodon interaction.

In a preferred embodiment, the invention permits the selection of nucleic acids by selecting the proteins for which they code. This may be accomplished by connecting the protein to its cognate mRNA at the end of translation, which in turn is done by connecting both the protein and mRNA to a tRNA or tRNA analog.

A preferred embodiment of the invention includes a tRNA molecule capable of covalently linking a nucleic acid encoding a polypeptide and the polypeptide to the tRNA, wherein the linkage of the nucleic acid occurs on a portion of the tRNA other than the linkage to the polypeptide and wherein the tRNA comprises a linking molecule associated with the anticodon of the tRNA. This anticodon of the tRNA is capable of forming a crosslink to the mRNA under irradiation with light of a required wavelength, preferably a furan-sided psoralen monoadduct on the anticodon irradiated with UVA, preferably in the range of about 300–450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. Preferably, an amino acid or amino acid analog is attached to the 3' end of a tRNA molecule by a stable bond to generate a stable aminoacyl tRNA analog (SATA).

Other embodiments include a mRNA comprising a psoralen, preferably located in the 3' region of the reading frame, more preferably at the most 3' codon of the reading frame, most preferably at the 3' stop codon of the reading frame. In preferred embodiments, linkage between the tRNA and the mRNA is a cross-linked psoralen molecule, more preferably a furan-sided psoralen monoadduct.

A further embodiment of the invention provides a method of forming a monoadduct. According to this method a target oligonucleotide with at least one uridine and at least one modified uridine is contacted with psoralen, and the target oligonucleotide and psoralen are coupled to form a monoadduct. The modified uridine

according to this embodiment may be modified to avoid coupling with psoralen, and preferably the modified uridine is pseudouridine. According to this embodiment the target oligonucleotide may be a tRNA molecule, such as tRNA, modified tRNA and tRNA analogs or a mRNA molecule, such as mRNA, modified mRNA and mRNA analogs. In a further embodiment the psoralen is coupled to the target oligonucleotide by one or more cross-links. According to this embodiment a second oligonucleotide with a nucleotide sequence complementary to the target oligonucleotide sequence may be present. This second oligonucleotide may contain no uridine or may contain uridine residues that are modified to avoid cross-linking with the target oligonucleotide. Preferably, the modified uridine is pseudouridine.

Several embodiments of the present invention include a method of stably linking a nucleic acid, a tRNA, and a polypeptide encoded by the nucleic acid together to form a linked nucleotide-polypeptide complex. In a preferred embodiment, the nucleic acid is an mRNA and the linked nucleotide-polypeptide complex is a mRNA-polypeptide complex. The method can further comprise providing a plurality of distinct nucleic acid-polypeptide complexes, providing a ligand with a desired binding characteristic, contacting the complexes with the ligand, removing unbound complexes, and recovering complexes bound to the ligand.

Several methods of the current invention involve the evolution of nucleic acid molecules and/or proteins. In one embodiment, this invention comprises amplifying the nucleic acid component of the recovered complexes and introducing variation to the sequence of the nucleic acids. In other embodiments, the method further comprises translating polypeptides from the amplified and varied nucleic acids, linking them together using tRNA, and contacting them with the ligand to select another new population of bound complexes. Several embodiments of the present invention use selected protein-mRNA complexes in a process of in vitro evolution, in particular the iterative process in which the selected mRNA is reproduced with variation, translated and again connected to cognate protein for selection.

In one embodiment, a strategy for selection is provided. In one embodiment, this strategy comprises the production of mRNA libraries. In one embodiment, RNA ligation is used. In one embodiment, RNA ligation using T4 RNA ligase is used. In

one embodiment, a diagnostic test for Severe Acute Respiratory Syndrome (SARS) is provided.

Several embodiments of the present invention provide compositions and methods for the efficient and rapid identification and selection of nucleic acids and polypeptides. Certain embodiments are particularly advantageous because the identification, selection, and/or evolution of nucleic acids and proteins according to one embodiment of the invention accommodates access to a large and highly varied population of test molecules, a way to select members of the population that exhibit the desired properties, and the ability to reproduce the selected molecules with mutated variations to obtain another large population for subsequent selection.

Several embodiments of the invention are useful for identifying and selecting genes and proteins used in the prevention and treatment of several diseases. For example, if a nucleic acid sequence linked to a disease is known, several embodiments of this invention can be used to quickly and accurately identify and select the corresponding protein. This protein can be then be mass-produced and used as diagnostic or therapeutic agents. Further, if an protein linked to a disease is known, several embodiments of this invention can permit the rapid identification of the corresponding nucleic acid. The nucleic acid can then be used as a diagnostic or therapeutic agent.

Another advantage of several embodiments of the present invention is the ability to overcome the inability of the proteins to reproduce themselves and the inability to link mRNA encoding a polypeptide with the translated product. Additionally, the generation of large peptide libraries and screening methods have, until recently, required that the process have an *in vivo* expression step. Examples include yeast two- or three-hybrid, yeast display and phage display methods (Fields and Song, *Nature* 340:245 (1989); Licitra and Liu, *PNAS* 93:12817 (1996); Boder and Wittrup, *Nat Biotechnol* 15:553 (1997); and Scott and Smith, *Science* 249:386 (1990)). *In vivo* methods, in some cases, suffer from disadvantages, including a limited library size and cumbersome screening steps. Additionally, undesirable selective pressures can be placed on the generation of variants by cellular constraints of the host. Notwithstanding the foregoing, one of skill in the art will appreciate that one embodiment of the current invention can be used using these *in vivo* methods.

*In vitro* methods have been developed more recently, using prokaryotic and eukaryotic *in vitro* translation systems, such as ribosome display (Mattheakis *et al.*, PNAS 91:9022 (1994); Hanes and Plückthun, PNAS 94:4937 (1997); Jermutus *et al.*, Current Opinion in Biotechnology 9:534 (1998), all herein incorporated by reference). These methods link the protein and its encoding mRNA with the ribosome, and the entire complex is screened against a ligand of choice. Potential disadvantages of this method include the large size of the ribosome, which could interfere with the screening of the attached, and relatively tiny, protein. One of skill in the art will appreciate that one embodiment of the current invention can be used using these *in vitro* methods.

In 1997, two groups of workers developed an *in vitro* method of attaching a protein to its coding sequence during translation by using the ribosomal peptidyl transferase with puromycin attached to a linker DNA (Szostak *et al.*, International Patent Publication WO 98/31700; Roberts and Szostak PNAS 94:12297 (1997); Nemoto *et al.*, FEBS Letters 414:405 (1997), all herein incorporated by reference). Once the coding sequence and peptides are linked, the peptides are exposed to a selected ligand. Selection or binding of the peptide by the ligand also selects the attached coding sequence, which can then be reproduced by standard means. Both Roberts and Szostak and Nemoto *et al.* used the technique of attaching a puromycin molecule to the 3' end of a coding sequence by a DNA linker or other non-translatable chain. Puromycin is a tRNA acceptor stem analog which accepts the nascent peptide chain under the action of the ribosomal peptidyl transferase and binds it stably and irreversibly, thereby halting translation.

Several embodiments of the current invention are particularly advantageous because they overcome one or more of the following limitations: (1) the coding sequence encoding each peptide must be known and be modified both initially and between each selection; (2) selection of native unknown mRNAs only; (3) the modification of the coding sequence adds several steps to the process; and (4) the attached puromycin on the linker molecules may compete in the translation reaction with the native tRNAs for the A site on the ribosome reading its coding sequence or a nearby ribosome, and could thus "poison" the translation process, just as would unattached puromycin in the translation reaction solution. Inadvertent interactions between puromycin and ribosomes could result in two kinds of reaction non-specificity:

prematurely shortened proteins and proteins attached to the wrong message. There are reports in the prior art that indicate that the avidity of the A site and the peptidyl transferase for the puromycin may be modulated by  $Mg^{++}$  concentration (Roberts, Curr. Opin. Chem. Biol. 3:268 (1999), herein incorporated by reference). Although  $Mg^{++}$  concentration may be titrated to control for the first kind of non-specificity (e.g., premature termination of translation), it will not affect the second type (e.g., inaccurate mRNA-protein linkage).

Other advantages of some embodiments of the present invention include the ability to generate high yield of cross-links, the ability to use a full complement of amino acids and the ability to use stop codons.

Thus, a need exists for an *in vitro* nucleic acid-based protein evolution system that, in some embodiments, does not necessarily require initial knowledge of the nucleic acid's sequence or repeated chemical modification of the nucleic acids, and which can accurately link a mRNA to its protein. There also remains a need for a system that is capable of using the full complement of amino acids with good efficiency in the presence of stop codons.

Several embodiments of the present invention provide compositions and methods to identify, select and evolve desired properties of proteins and nucleic acids. In many embodiments, the current invention provides tRNA molecules, which include modified tRNAs and tRNA analogs. In other embodiments, tRNA molecules include native or unmodified tRNAs. Other embodiments include methods for generating polypeptides, assays enabling selection of individual members of a population of polypeptides having desired characteristics, methods for amplifying the nucleic acids encoding such selected polypeptides, and methods for generating new variants to screen for enhanced properties.

In several embodiments, the present invention permits the attachment of a protein to its respective mRNA without requiring modification of native mRNA. In another embodiment, a vaccine for SARS, and a method for making same, are provided. Only minimal modification is needed. In yet another embodiment, extensively modified mRNA can be used. The specificity of the methods embodied in some embodiments are determined by the specificity of the codon-anticodon interaction.

In a preferred embodiment, the invention permits the selection of nucleic acids by selecting the proteins for which they code. This, in one embodiment, this is accomplished by connecting the protein to its cognate mRNA at the end of translation, which in turn is done by connecting both the protein and mRNA to a tRNA molecule.

In one embodiment, a method for identifying a desired protein or nucleic acid molecule is provided. In one embodiment, at least two mRNA molecules are provided. At least one of the mRNA molecules comprises a stop codon and/or a pseudo stop codon. The mRNA molecules is translated to generate at least one translated protein. The mRNA molecules is linked, coupled or associated to its corresponding translated protein using a tRNA molecule to form at least one cognate pair. At least one of the mRNA molecules is connected to the tRNA molecule by a crosslinker. In one embodiment, the cognate pairs is identified using a property of the translated protein or the mRNA molecule. An mRNA molecule of the selected cognate pair, a nucleic acid molecule complementary to the mRNA molecule and/or a nucleic acid molecule homologous to the mRNA molecule is identified, thereby identifying the desired protein or the desired nucleic acid molecule.

In one embodiment, the tRNA molecule is a stable aminoacyl tRNA analog (SATA). As used herein, a SATA is an entity which can recognize a selected codon such that it can accept a peptide chain by the action of the ribosomal peptidyl transferase, preferably when the cognate codon is in the reading position of the ribosome.

In one embodiment, the SATA comprises a puromycin and a crosslinker that are both located on the SATA. The term "located on" as used herein shall be given its ordinary meaning and shall also meaning positioned on, incorporated in, attached to, coupled to, bound to, or integral to. In one embodiment, the SATA comprises a puromycin, but the crosslinker is located on the mRNA molecule. In one embodiment, the crosslinker is located only on the mRNA and not on the tRNA.

In one embodiment, the tRNA molecule is a Linking tRNA Analog. In one embodiment, a crosslinker is located on the Linking tRNA Analog, and no puromycin is present.

In one embodiment, the tRNA molecule is a Nonsense Suppressor tRNA. In one embodiment, a crosslinker is located not on the tRNA, but on the mRNA, and no

puromycin is present. In one embodiment, the crosslinker is located only on the mRNA and not on the tRNA. In one embodiment, the Nonsense Suppressor tRNA is a substantially unmodified native tRNA.

In one embodiment of the invention, the crosslinker is an agent that chemically or mechanically links two molecules together. In one embodiment, the crosslinker is an agent that can be activated to form one or more covalent bonds with tRNA and/or mRNA. In one embodiment, the crosslinker is a sulfur-substituted nucleotide. In another embodiment, the crosslinker is a halogen-substituted nucleotide. Examples of crosslinkers include, but are not limited to, 2-thiocytosine, 2-thiouridine, 4-thiouridine, 5-iodocytosine, 5-iodouridine, 5-bromouridine and 2-chloroadenosine, aryl azides, and modifications or analogues thereof. In one embodiment, the crosslinker is psoralen or a psoralen analog. One or more crosslinkers can be used, and the locations of these crosslinkers can be varied.

In one embodiment, the crosslinker is located on the mRNA. In another embodiment, the crosslinker is located on the tRNA molecule. In one embodiment, the crosslinker is located on or near a codon. In another embodiment, the crosslinker is located on or near a stop or pseudo stop codon. In one embodiment, the crosslinker is located on or near an anticodon of the RNA molecule. In one embodiment, the crosslinker is located on or near a stop or pseudo stop anticodon of the RNA molecule.

In one embodiment of the invention, the crosslinker forms a bond or coupling between the tRNA molecule and the mRNA molecule. In one embodiment, the tRNA molecule is connected to its translated protein by ribosomal peptidyl transferase. In another embodiment, the tRNA molecule is connected to the mRNA through an ultraviolet-induced crosslink between the anticodon of the tRNA molecule and the codon of the mRNA.

In one embodiment, the tRNA molecule has a stable peptide acceptor. The stable peptide acceptor, in one embodiment, is a puromycin or puromycin analog. In one embodiment, the tRNA molecule is operable to accept a peptide chain and hold the chain in a stable manner such that ribosomal peptidyl transferase cannot detach it. In one embodiment, the tRNA molecule comprises a moiety which binds to the ribosome, accepts the peptide chain, and then does not act as a donor in the next transpeptidation. The moiety can be located on the tRNA. In one embodiment, the moiety includes, but is



not limited to, a 2' ester on a 3' deoxy adenosine, an amino acyl tRNAox-red and a puromycin. One or more moieties may be located on the tRNA molecule.

In one embodiment, the mRNA molecule is untranslatable beyond a linking codon. In one embodiment, the tRNA molecule accepts a peptide chain and holds the chain in a manner such that ribosomal peptidyl transferase cannot detach it because the message in subsequent codons is untranslatable. In another embodiment, the tRNA molecule accepts a peptide chain and holds the chain in a manner such that ribosomal peptidyl transferase cannot detach it because the message is untranslatable. The message can be untranslatable because it is at the end of the message or because the tRNAs that recognize the appropriate codons have been depleted. Other techniques to make the mRNA untranslatable can also be used.

In one embodiment of the current invention, translation is performed *in vitro*. In another embodiment, translation is performed *in situ*. In yet another embodiment, *in vivo* translation is provided.

In another embodiment of the invention, the method further comprises selecting a desired nucleic acid or protein by providing a plurality of cognate pairs, binding at least one of these cognate pairs with one or more binding agents, and selecting the desired protein or nucleic acid molecule based upon a reaction to the binding agents. Selection can also be performed based on a lack of reaction to a binding agent.

In one embodiment, the step of providing a plurality of cognate pairs comprises providing one or more cognate pairs on or in a medium selected from the group consisting of one or more of the following: a matrix, in solution, on beads, and on an array. One skilled in the art will understand that cognate pairs can be placed in any medium suitable for further binding or selection. In one embodiment, the cognate pair is selected based upon ligand binding. Ligands include, but are not limited to, proteins, nucleic acids, chemical compounds, polymers and metals. In another embodiment, the reaction is selected from the group consisting of one or more of the following: ligand binding, immunoprecipitation, and enzymatic reactions. One skilled in the art will understand that any reaction that serves to distinguish the target molecule can be used. These reactions include, but are not limited to, chemical, mechanical, and biological reactions.

In another embodiment of the invention, the method further comprises selecting a desired nucleic acid molecule. In one embodiment, the method comprises providing an array of nucleic acids, wherein the nucleic acids are placed in a predetermined position, hybridizing at least one of the cognate pairs onto the array, reacting the cognate pairs with one or more binding agents, and selecting the desired nucleic acid molecule based upon a reaction or lack of a reaction to the binding agent. Binding agents include, but are not limited to ligands, described above. One skilled in the art will understand that any reaction that serves to distinguish the desired nucleic acid molecule can be used. These reactions include, but are not limited to, chemical, mechanical, and biological reactions.

In yet another embodiment, the method further comprises determining the DNA sequence of the translated protein. In one embodiment, the method comprises providing an array of two or more DNA sequences, wherein the DNA sequences are placed in a predetermined position, exposing the array to one or more cognate pairs, wherein one or more cognate pairs comprises an mRNA portion and a protein portion, hybridizing the mRNA portion of the cognate pairs onto the array, exposing the protein portion of one or more cognate pairs to a binding agent, thereby producing a reaction or a non-reaction, and selecting the desired protein based upon the reaction or non-reaction to the binding agent, such as a ligand, thereby determining the DNA sequence of the translated protein.

In one embodiment of the present invention, a modified mRNA molecule operable to crosslink to a tRNA molecule is provided. In one embodiment, the modified mRNA molecule comprises a crosslinker located on or near a stop codon. In one embodiment, the modified mRNA molecule comprises a crosslinker located on or near a pseudo stop codon.

In one embodiment, the crosslinker is an agent that can be activated to form one or more covalent bonds with the tRNA. In one embodiment, the crosslinker is an agent that is activated to form one or more covalent bonds with the tRNA using light. In another embodiment, the crosslinker is a modified base that is incorporated directly into the mRNA. In one embodiment, crosslinker is selected from the group consisting of one or more of the following 2-thiocytosine, 2-thiouridine, 4-thiouridine, 5-iodocytosine, 5-iodouridine, 5-bromouridine and 2-chloroadenosine, aryl azides, and

modifications or analogues thereof. In several embodiments, the crosslinker is psoralen.

In one embodiment of the present invention, a kit to generate cognate pairs is provided. In one embodiment, the kit is a compilation, collection, system or group of items that comprise at least one psoralen monoadduct attached to a nonadducted stable aminoacyl tRNA analog. In another embodiment, the kit comprises at least one psoralen monoadduct attached to an oligonucleotide. In several embodiments, the kit comprises instructions regarding the generation of cognate pairs. In yet another embodiment, the kit comprises additional chemicals, agents or equipment that would be useful to generate cognate pairs.

In one embodiment of the invention, a method for evolving desired sequences is provided. In one embodiment, the method comprises: providing at least two candidate mRNA molecules, wherein the mRNA molecule contains a stop codon and/or a pseudo stop codon; translating at least two of the mRNA molecules to generate at least one translated protein, linking at least one of the mRNA molecules to its corresponding translated protein via a tRNA molecule to form at least one cognate pair, wherein at least one of the candidate mRNA molecules is connected to the tRNA molecule by a crosslinker, identifying one or more of the cognate pairs based upon the properties of the translated protein or the mRNA molecule, identifying a molecule selected from the group consisting of one or more of the following: an mRNA molecule of the selected cognate pair, a nucleic acid molecule complementary to the mRNA molecule and a nucleic acid molecule homologous to the mRNA molecule, thereby identifying the desired protein or the desired nucleic acid molecule. The method, in some embodiments, further comprises providing a plurality of cognate pairs, binding at least one of the plurality of cognate pairs with one or more binding agents, selecting the desired protein or nucleic acid molecule based upon a reaction or lack of a reaction to the one or more binding agents, thereby selecting a first desired cognate pair. The method, in several embodiments, further comprises recovering the first desired cognate pair to generate a recovered cognate pair, amplifying a first nucleic acid component of the recovered cognate pair, producing a second nucleic acid component, wherein the second nucleic acid component comprises the first nucleic acid component with one or more variations, producing a second protein by translating the second nucleic acid

component, linking the second protein with the second nucleic acid component to generate a second desired cognate pair, and obtaining the desired protein sequence by re-selecting the second desired cognate pair based upon at least one desired property. In a preferred embodiment, the desired sequence is a sequence for is one or more sequences for the SARS virus.

In one embodiment, the desired property is selected from the group consisting of one or more of the following: binding properties, enzymatic reactions and chemical modifications. In one embodiment, the desired property is a lack of a reaction (or an ability to resist binding, enzymatic reaction or chemical modification). In one embodiment, the step of selecting the first desired cognate pair comprises: providing a first ligand with a desired binding characteristic, contacting one or more of the first cognate pairs with the first ligand to generate unbound complexes and bound complexes, recovering either the bound complexes or the unbound complexes, amplifying at least one nucleic acid component of the recovered complexes, introducing variation to a sequence of the nucleic acid component of the recovered complexes, translating one or more second proteins from the nucleic acid components, linking at least one of the second proteins with at least one of the second nucleic acid components to generate one or more second cognate pairs, and obtaining the desired protein sequence by contacting the at least one of the second cognate pairs with at least one second ligand to select one or more of the second cognate pairs, wherein the second ligand is the same or different than the first ligand.

One embodiment of the invention is directed to a method of forming a monoadduct which includes the steps of providing a target oligonucleotide including at least one uridine and at least one modified uridine, contacting said target oligonucleotide with psoralen, and coupling said psoralen to said target oligonucleotide to form a monoadduct.

One embodiment of the invention is directed to a method for identifying and selecting a desired protein or nucleic acid molecule including the steps of providing at least two candidate mRNA molecules, wherein at least one of said mRNA molecules contains at least one codon which is a stop codon or a pseudo stop codon; translating at least two of said candidate mRNA molecules to generate at least one translated protein; linking at least one of said candidate mRNA molecules to its corresponding translated

protein via a tRNA molecule to form at least one cognate pair, wherein at least one of said candidate mRNA molecules is connected to said tRNA molecule by a crosslinker; identifying one or more of said cognate pairs based upon the properties of said translated protein or said mRNA molecule; identifying a molecule which is one or more of the following: an mRNA molecule of said selected cognate pair, a nucleic acid molecule complementary to said mRNA molecule and a nucleic acid molecule homologous to said mRNA molecule, thereby identifying said desired protein or said desired nucleic acid molecule; providing a plurality of cognate pairs, binding at said plurality of cognate pairs with one or more binding agents; and selecting said desired protein or nucleic acid molecule based upon a reaction or lack of a reaction to said one or more binding agents.

In one embodiment of the present invention, the invention comprises a method of forming a psoralen monoadduct on a nucleic acid. In one embodiment, the method comprises providing a first nucleic acid and a second nucleic acid, wherein the first nucleic acid and the second nucleic acid are substantially complementary to each other, wherein the first nucleic acid comprises one or more uridine monoadduct targets, and wherein the second nucleic acid comprises at least one pseudouridine. The method further comprises hybridizing at least a portion of the first nucleic acid and the second nucleic acid in the presence of psoralen to form a hybrid, irradiating the hybrid with ultraviolet light, thereby forming the psoralen monoadduct on the first nucleic acid. In one embodiment, one or more uridine monoadduct targets comprises a uridine located adjacent to an adenosine, preferably 3' from the adenosine.

In one embodiment of the invention, a method of producing a psoralen monoadduct or a crosslink is provided. In one embodiment, the method comprises providing a first nucleic acid and a second nucleic acid, wherein the first nucleic acid and the second nucleic acid are substantially complementary to each other, wherein the first nucleic acid comprises one or more uridine monoadduct targets or crosslink targets and one or more uridine monoadduct non-targets or crosslink non-targets, and wherein the uridine monoadduct non-targets or crosslink non-targets are operable to be replaced with one or more pseudouridines. The method further comprises replacing one or more of the uridine monoadduct non-targets or crosslink non-targets with pseudouridine, hybridizing at least a portion of the first nucleic acid and the second nucleic acid in the

presence of psoralen to form a hybrid; and irradiating the hybrid, thereby forming the psoralen monoadduct or the crosslink on the first nucleic acid on the targets, while protecting the nontargets. In one embodiment, visible light is used to form the adduct or crosslink. In another embodiment, ultraviolet light is used.

One embodiment of the invention is directed to a method for evolving a desired protein sequence which includes the steps of providing at least two candidate mRNA molecules, wherein at least one of said mRNA molecules contains at least one codon which is a stop codon or a pseudo stop codon; translating at least two of said candidate mRNA molecules to generate at least one translated protein; linking at least one of said candidate mRNA molecules to its corresponding translated protein via a tRNA molecule to form at least one cognate pair, wherein at least one of said candidate mRNA molecules is connected to said tRNA molecule by a crosslinker; identifying one or more of said cognate pairs based upon the properties of said translated protein or said mRNA molecule; identifying a molecule selected from one or more of the following: an mRNA molecule of said selected cognate pair, a nucleic acid molecule complementary to said mRNA molecule and a nucleic acid molecule homologous to said mRNA molecule, thereby identifying said desired protein or said desired nucleic acid molecule; providing a plurality of cognate pairs, binding at least of said plurality of cognate pairs with one or more binding agents; selecting said desired or protein nucleic acid molecule based upon a reaction or lack of a reaction to said one or more binding agents, thereby selecting a first desired cognate pair; recovering said first desired cognate pair to generate a recovered cognate pair; amplifying a first nucleic acid component of said recovered cognate pair; producing a second nucleic acid component, wherein said second nucleic acid component comprises said first nucleic acid component with one or more variations; producing a second protein by translating said second nucleic acid component; linking said second protein with said second nucleic acid component to generate a second desired cognate pair; and obtaining the desired protein sequence by re-selecting said second desired cognate pair based upon at least one desired property.

In preferred embodiments, the step of selecting said first desired cognate pair includes the steps of providing a first ligand with a desired binding characteristic; contacting one or more of said first cognate pairs with said first ligand to generate

unbound complexes and bound complexes; recovering either the bound complexes or the unbound complexes; amplifying at least one nucleic acid component of the recovered complexes; introducing variation to a sequence of said nucleic acid component of said recovered complexes; translating one or more second proteins from said nucleic acid components, linking at least one of said second proteins with at least one of said second nucleic acid components to generate one or more second cognate pairs; and obtaining the desired protein sequence by contacting said at least one of said second cognate pairs with at least one second ligand to select one or more of said second cognate pairs, wherein said second ligand is the same or different than said first ligand.

Some embodiments are directed to a method of forming a psoralen monoadduct on a nucleic acid, including the steps of providing a first nucleic acid and a second nucleic acid, wherein said first nucleic acid and said second nucleic acid are substantially complementary to each other, wherein said first nucleic acid comprises one or more uridine monoadduct targets, and wherein said second nucleic acid comprises at least one pseudouridine hybridizing said first nucleic acid and said second nucleic acid in the presence of psoralen to form a hybrid; irradiating said hybrid with ultraviolet light, thereby forming said psoralen monoadduct on said first nucleic acid.

Some embodiments are directed to a method of producing a psoralen monoadduct or a crosslink, including the steps of providing a first nucleic acid and a second nucleic acid; wherein said first nucleic acid and said second nucleic acid are substantially complementary to each other; wherein said first nucleic acid includes one or more uridine monoadduct targets or crosslink targets and one or more uridine monoadduct non-targets or crosslink non-targets; wherein said uridine monoadduct non-targets or crosslink non-targets are operable to be replaced with one or more pseudouridines; replacing one or more of said uridine monoadduct non-targets or crosslink non-targets with pseudouridine; hybridizing said first nucleic acid and said second nucleic acid in the presence of psoralen to form a hybrid; irradiating said hybrid, thereby forming said psoralen monoadduct or said crosslink on said first nucleic acid on said targets, while protecting said nontargets.

Several embodiments of the present invention are directed to vaccine production. In one embodiment, rather than select for a few proteins with the highest binding

affinities in a given distribution, a less stringent selection is used so as to have a high number of different sequences and use multiple rounds of mutation with gradual increase in the stringency to evolve a large population of proteins with a high binding affinity. Such proteins are of value for making vaccines. The logic is similar to an anti-idiotypic vaccine except that there will be one and only one surface epitope that can react with the immune system. The aggregate concentration of the desired protein presented to the immune system by the family of proteins will be sufficiently high to reach the threshold level required to stimulate a T-cell and B-cell response. However, the concentration of any single protein within the family will be below the threshold required to stimulate a response to that protein. Therefore, the vaccine will stimulate antibody production only against the desired epitope and not against any of the other epitopes present on the family of proteins. This will prevent production of antibodies that could inactivate the vaccine. In another embodiment, the vaccine will be synthesized such that it will stimulate antibody production against the desired epitope and one or more other epitopes that have either a neutral or synergistic effect with activation of the desired epitope.

#### Brief Description of the Drawings

Figure 1 illustrates schematically one example of the complex formed by the mRNA and its protein product when linked by a modified tRNA or analog. As shown, a codon of the mRNA pairs with the anticodon of a modified tRNA and is covalently crosslinked to a psoralen monoadduct, or a non-psoralen crosslinker or aryl azides by UV irradiation. The translated polypeptide is linked to the modified tRNA via the ribosomal peptidyl transferase. Both linkages occur while the mRNA and nascent protein are held in place by the ribosome.

Figure 2 illustrates schematically an example of the *in vitro* selection and evolution process, wherein the starting nucleic acids and their protein products are linked (e.g., according to Figure 1) and are selected by a particular characteristic exhibited by the protein. Proteins not exhibiting the particular characteristic are discarded and those having the characteristic are amplified with variation, preferably via amplification with variation of the mRNA, to form a new population. In various embodiments, nonbinding proteins will be selected. The new population is translated and linked via a modified tRNA or analog, and the selection process is repeated. As



many selection and amplification/mutation rounds as desired can be performed to optimize the protein product.

Figure 3 illustrates one method of construction of a tRNA molecule of the invention. In this embodiment, the 5' end of a tRNA, a nucleic acid encoding an anticodon loop and having a molecule capable of stably linking to mRNA (such as psoralen, as used in this example), and the 3' end of tRNA modified with a terminal puromycin molecule are ligated to form a complete modified tRNA for use in the *in vitro* evolution methods of the invention. Other embodiments do not include puromycin.

Figure 4 describes two alternative embodiments by which the crosslinking molecule psoralen can be positioned such that it is capable of linking the mRNA with the tRNA in the methods of the invention. A first embodiment includes linking the crosslinker (e.g., psoralen monoadduct) to the mRNA, and a second embodiment includes linking the crosslinker to the anticodon of the tRNA molecule. The crosslinker can either be monoadducted to the anticodon or the 3' terminal codon of the reading frame for known or partially known messages. This can be done in a separate procedure from translation, e.g., before translation occurs.

Figure 5 illustrates the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does, but lacks the 5-6 double bond that is the target for psoralen.

Figure 6 illustrates some embodiments of the present invention. The SATA, Linking tRNA Analog and Nonsense Suppressor analog, in certain embodiments, are shown.

Figure 7 shows the probability of obtaining a nucleotide of a given length.

Figure 8 shows a reaction scheme for producing an mRNA for a protein of 128 amino acids.

Figure 9 shows conc pT=6 vs. log k.

Figure 10 shows concentration vs. log k

Figure 11 shows Lancet's empiric distribution vs. a Poisson with the same mean.

Figure 12 shows family of curves that would be bound by different [T] values.

Figure 13 shows 4 generations with  $[T] = 10^{-12}$

Figure 14 shows comparison of normal protein synthesis and non-limiting embodiments of methods for protein synthesis. Normal Translation: After translation in the ribosome, the mRNA and the resultant protein become separated. Preferred embodiment (A) shows the preferred linker is on a SATA linking reagent and, after in vitro translation, the mRNA and the protein are linked by the SATA linker. Preferred embodiment (B) shows the preferred linker is on the mRNA and, after in vitro translation, the mRNA and the protein are connected by the linker located on the mRNA.

Figure 15 shows an illustration of how one non-limiting embodiment of the linker technology, coupled with one embodiment of its proprietary method for making a starting mRNA library, and one embodiment of its proprietary selection procedure, enables one to rapidly isolate the mRNA from a protein of interest and then use it to generate production scale amounts of that protein or to accelerate the creation, and then selection of, proteins with enhanced properties.

Figure 16 shows a random library linked with a SARS "S" protein or other target "T" protein (henceforth "ST") protein coated Surface Plasmon Resonance (SPR) membrane to generate the distribution of binding constants for the protein library..

Figure 17 shows "S" or "T" protein on SPR membrane shown with all of the Ptrap binding domains saturated with trapping protein (Ptrap) and showing signaling protein (Psig) bound to a different domain.

Figure 18 shows polyacrylamide magnetic bead coated with trapping probe .

Figure 19 shows gold particle coated with psig protein and with bar code oligonucleotides.

Figure 20 shows assay for SARS virus.

Figure 21 shows Protein-SATA-mRNA library captured on SPR membrane by anti-S antibody.

Figure 22 shows cancer treatment.

#### Detailed Description of the Preferred Embodiment

Various aspects of the present invention use a tRNA mechanism that links messenger RNA (mRNA) to its translated protein product, forming a "cognate pair." In several embodiments, an mRNA, whose sequence is not known, can be expressed, its protein characterized through a selection process against a ligand with desired or

selected properties, and nucleic acid evolution—resulting in protein evolution—can be performed *in vitro* to arrive at molecules with enhanced properties. The cognate pairs are preferably attached via a tRNA molecule.

The term “tRNA molecule”, as used herein, shall be given its ordinary meaning and shall also mean a stable aminoacyl tRNA analog (SATA), a Linking tRNA Analog, and a Nonsense Suppressor Analog, all of which are described herein. A tRNA molecule includes native tRNA, synthetic tRNA, a combination of native and synthetic tRNA, and any modifications thereof. In a preferred embodiment, the tRNA is connected to the nascent peptide by the ribosomal peptidyl transferase and to the mRNA through an ultraviolet induced crosslink between the anticodon of the tRNA molecule and the codon of the RNA message. This can be done by, for example, thiouracil. In one preferred embodiment, the linker is a psoralen crosslink made from a psoralen monoadduct, a non-psoralen crosslinker, or analogs or modifications thereof, pre-placed on either the mRNA's last translatable codon or preferably on the tRNA anticodon of choice. Preferably, a tRNA stop anticodon is selected. A stop codon/anticodon pair selects for full length transcripts. One skilled in the art will understand that an mRNA not having a stop codon may also be used and, further, that any codon or nucleic acid triplet may be used in accordance with several embodiments of the current invention. A tRNA having an anticodon which is not naturally occurring can be synthesized according to methods known in the art (e.g. Figure 3).

In one embodiment, the anticodon of the tRNA is capable of forming a crosslink to the mRNA, where the cross-link is selected from the group consisting of one or more of the following: 2-thiocytosine, 2-thiouridine, 4-thiouridine, 5-iodocytosine, 5-iodouridine, 5-bromouridine and 2-chloroadenosine, aryl azides, and modifications or analogues thereof. These crosslinkers are available commercially from Ambion, Inc. (Austin, TX), Dharmacon, Inc. (Lafayette, CO), and other well-known manufacturers of scientific materials.

The terms “protein,” “peptide,” and “polypeptide” are defined herein to mean a polymeric molecule of two or more units comprised of amino acids in any form (e.g., D- or L- amino acids, synthetic or modified amino acids capable of polymerizing via peptide bonds, etc.), and these terms may be used interchangeably herein.

The term “pseudo stop codon” is defined herein to mean a codon which, while not naturally a nonsense codon, prevents a message from being further translated. A pseudo stop codon may be created by using a “stable aminoacyl tRNA analog” or SATA, as described below. In this manner, a pseudo stop codon is a codon which is recognized by and binds to a SATA. Another method by which to create a pseudo stop codon is to create an artificial system in which the necessary tRNA having an anticodon complementary to the pseudocodon is substantially depleted. Accordingly, translation will stop when the absent tRNA is required, e.g., at the pseudo stop codon.

In another embodiment, the selected codon is located on, or placed at, the end of the translatable reading frame by one or more of the following methods (1) having it be the 3' end; (2) providing or having modifications to the moieties 3' to the linking codon, thereby rendering them untranslatable and incapable of activating release factors; and (3) by having codons 3' to the linking codon whose corresponding tRNAs have been depleted.

One skilled in the art will appreciate that there are several ways to create a pseudo stop codon that can be used in accordance with several embodiments of the present invention.

The formation of connections between mRNA and its protein product generally requires a tRNA, tRNA analog, or an mRNA with certain characteristics. In several embodiments of the current invention, the tRNA or tRNA analog will have a stable peptide acceptor. This modification changes the tRNA or tRNA analog such that after it accepts the nascent peptide chain by the action of the ribosomal peptidyl transferase, it holds the chain in a stable manner such that the peptidyl transferase cannot detach it. This may be accomplished by using a bond such as a 2' ester on a 3' deoxy adenosine or an amino “acyl tRNA<sub>ox-red</sub>” which can bind to the ribosome, accept the peptide chain, and then not act as a donor in the next transpeptidation (Chinali *et al.*, *Biochem.* 13:3001 (1974); Krayevsky and Kukhanova, *Prog. Nuc. Acid Res* 23:1 (1979) and Sprinzl and Cramer *Prog. Nuc. Acid Res* 22:1 (1979), all herein incorporated by reference).

In a further embodiment, a selected codon is located on, or placed at, the end of the translatable reading frame by having it be the 3' end or by providing modifications

to the more 3' moieties rendering them untranslatable and incapable of recognizing release factors.

In one embodiment, an amino acid or amino acid analog is attached to the 3' end of the tRNA or tRNA analog by a stable bond. This stable bond contrasts the labile, high energy ester bond that connects these two in the native structure. The stable bond not only protects the bond from the action of the peptidyl transferase, but also preserves the structure during subsequent steps. For convenience, this modified tRNA or tRNA analog will be referred to as a "stable aminoacyl tRNA analog" or SATA. As used herein, a SATA is an entity which can recognize a selected codon such that it can accept a peptide chain by the action of the ribosomal peptidyl transferase when the cognate codon is preferably in the reading position of the ribosome. The peptide chain will be bound in such a way that the peptide is bound stably and cannot be unattached by the peptidyl transferase. Preferably, the selected codon is recognized by hydrogen bonding.

One method for creating a stabilized modified tRNA was published in 1973 (Fraser and Rich, PNAS 70:2671 (1973), herein incorporated by reference). This method involves the conversion of a tRNA, or tRNA analog, to a 3'-amino-3'-deoxy tRNA. This is accomplished by adding a 3'-amino-3'-deoxy adenosine to the end of a native tRNA with tRNA nucleotidyl transferase after removing the native adenosine from it with snake venom phosphodiesterase. This modified tRNA is then charged with an amino acid by the respective aminoacyl tRNA synthetase (aaRS). Fraser and Rich used an aaRS in which the tRNA is charged on the 3', rather than the 2', hydroxyl. The amino acid is bound to the tRNA by a stable amide bond rather than the usual labile high-energy ester bond. Thus, when it accepts a peptide from ribosomal peptidyl transferase it will stably hold the peptide and not be able to donate it to another acceptor.

In a preferred method, the SATA will be attached to the translated message by a psoralen cross link between the codon and anticodon. Psoralen cross links are preferentially made between sequences that contain complementary 5' pyrimidine-purine 3' sequences especially UA or TA sequences (Cimino et al., Ann. Rev. Biochem. 54:1151 (1985), herein incorporated by reference). The codon coding for the SATA, or the linking codon, can be PYR-PUR-X or X-PYR-PUR, so that several codons may be

used for the linking codon. Conveniently, the stop or nonsense codons have this configuration. Using a codon that codes for an amino acid may require minor adjustments to the genetic code, which could complicate some applications. Therefore, in a preferred embodiment, a stop codon is used as the linking codon and the SATA functions as a nonsense suppressor in that it recognizes the linking codon. One skilled in the art, however, will appreciate that, with appropriate adjustments to the system, any codon can be used.

Fraser and Rich did their work in *E. coli*, but the most effective in vitro translation systems are in eukaryotes. The use of prokaryotic suppressors in eukaryotic translation systems appears to be feasible (Geller and Rich *Nature* 283:41 (1980); Edwards et al *PNAS* 88:1153 (1991); Hou and Schimmel *Biochem* 28:6800 (1989), all herein incorporated by reference). They are primarily limited by the resident aaRS's. This limitation is overcome by various embodiments of the present invention because the tRNA or analog can be charged in the prokaryotic system and then purified according to established methods (Lucas-Lenard and Haenni, *PNAS* 63:93 (1969), herein incorporated by reference).

In several embodiments of the current invention, acceptor stem modifications suitable for use in the tRNAs and analogs can be produced by various methods known in the art. Such methods are found in, for example, Sprinzl and Cramer, *Prog. Nuc. Acid Res.* 22:1 (1979), herein incorporated by reference. In an alternative embodiment, "transcriptional tRNA", i.e. the sequence of the tRNA as it would be transcribed rather than after the post-transcriptional processing, leads to the atypical and modified bases that are common in tRNAs. These transcriptional tRNAs are capable of functioning as tRNAs (Dabrowski et al., *EMBO J.* 14: 4872, 1995; and Harrington et al., *Biochem.* 32: 7617, 1993, both herein incorporated by reference). Transcriptional tRNA can be produced by transcription or can be made by connecting commercial RNA sequences together, piece-wise as in Figure 3, or in some combination of established methods. For instance, the 5' phosphate and 3' puromycin are commercially available attached to oligoribonucleotides. Commercial RNA sequences are available from Dharmacon Research Inc., La Fayette, CO. This company can also provide modified native tRNA, such as sequences in which thymine is substituted for uracil and pseudouridine.) These pieces can be connected together using T4 DNA ligase, as is well-known in the art

(Moore and Sharp, *Science* 256: 992, 1992, herein incorporated by reference). Alternatively, in a preferred embodiment, T4 RNA ligase is used (Romaniuk and Uhlenbeck, *Methods in Enzymology* 100:52 (1983), herein incorporated by reference).

In several embodiments of the present invention, psoralen is monoadducted to the SATA by construction of a tRNA from pieces including a psoralen linked oligonucleotide (Fig. 3) or by monoadduction to a native or modified tRNA or analog (Fig. 4). In a preferred embodiment, psoralen is first monoadducted to an oligonucleotide containing part of the anticodon loop as described below and this product is then ligated to the remaining fragments of the SATA.

In several embodiments, translation will stop when the nascent protein is attached to the SATA by the peptidyl transferase. When a large number of ribosomes are in this position the SATA and the mRNA will be connected with UV light. In a preferred method this will be accomplished by having a psoralen crosslink formed. Psoralens have a furan side and a pyrone side, and they readily intercalate between complementary base pairs in double stranded DNA, RNA, and DNA-RNA hybrids (Cimino et al., *Ann. Rev. Biochem.* 54:1151 (1985), herein incorporated by reference). Upon irradiation with UV, preferably in the range of 320 nm to 400 nm, cross linking will take place and leave the staggered pyrimidines covalently bound. By either forming crosslinks and photo reversing them or by using selected wavelengths, it is possible to form monoadducts, described more fully below. These will be either pyrone sided or furan sided monoadducts. Upon further irradiation, the furan sided monoadducts can be covalently crosslinked to complementary base pairs. The pyrone sided monoadducts cannot be further crosslinked. The formation of the furan sided psoralen monoadduct (MAf) is also done according to established methods. In a preferred method, the psoralen is attached to the anticodon of the SATA. However, psoralen can also be attached at the end of the reading frame of the message, as depicted in Figure 4.

Methods for large scale production of purified MAf on oligonucleotides are described in the literature (e.g., Speilmann et al., *PNAS* 89:4514, 1992, herein incorporated by reference), as are methods that require less resources, but have some non-cross-linkable pyrone sided psoralen monoadduct contamination (e.g., U.S. Patent No. 4,599,303; Gamper et al., *J. Mol. Biol.* 197: 349 (1987); Gamper et al., *Photochem.*

Photobiol. 40:29 (1984), both herein incorporated by reference). In several embodiments of the current invention, psoralen labeling is accomplished by using either method. In a preferred embodiment, furan sided monoadducts will be created using visible light, preferably in the range of approximately 400 nm - 420 nm, according to the methods described in U.S. Patent No. 5,462,733 and Gasparro et al., Photochem. Photobiol. 57:1007 (1993), both herein incorporated by reference. In one aspect of this invention, a SATA with a furan sided monoadduct or monoadducted oligonucleotides for placement on the 3' end of mRNAs, along with a nonadducted SATA are provided as the basis of a kit.

In one embodiment, the formation and reversal of monoadducts and crosslinks are performed according to the methods of Bachellerie et al. (Nuc Acids Res 9:2207 (1981)), herein incorporated by reference. In a preferred embodiment, efficient production of monoadducts, resulting in high yield of the end-product, is accomplished using the methods of Kobertz and Essigmann, J. A. Chem. Soc. 1997, 119, 5960-5961 and Kobertz and Essigmann, J. Org. Chem. 1997, 62, 2630-2632, both herein incorporated by reference.

In a preferred embodiment, a SATA fragment and complementary RNA or DNA is used in which all of the uridines, except the target, are replaced by pseudouridine. Figure 5 compares the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does. This embodiment is particularly advantageous because the pseudouridine forms the same Watson-Crick hydrogen-bonds as the native uridine but lacks the 5-6 double bond that is the target for interacting with either the furan or pyrone side of the psoralen molecule. This permits the same base-pairing characteristics as an oligonucleotide with uridine, but provides only one target for the psoralen. Because the pyrone side linkage is usually formed after the furan side has reacted, this removal of a staggered target allows the monoadduct to be formed with high efficiency irradiation without forming crosslinks and with minimal formation of pyrone sided monoadduct (MaP). Irradiation is preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. More specifically, a pseudouridine on the SATA permits: 1) the use of SATA sequences that contain uridines which are potential targets for the psoralen and



2) on the cRNA or cDNA, eliminate the formation of crosslinks, leaving the process stopped at furan sided monoadduct (MaF) formation when using UVA wavelengths which are much more efficient than visible light.

As described herein, non-psoralen crosslinkers, or modifications and analogues thereof, are used in several embodiments. One advantage of non-psoralen crosslinkers is that they are easier to work with in some instances because they can be incorporated into the tRNA or mRNA by commercially available means. For example, use of aryl azide is demonstrated in Demeshkina, N, et al RNA 6:1727-1736, 2000, herein incorporated by reference.

Use of the SATA and the monoadduct in several embodiments of the current invention is particularly advantageous for *in vitro* translation systems. However, one skilled in the art will appreciate that *in situ* systems can also be used. Various embodiments of the current invention will be applicable to any *in vitro* translation system, including, but not limited to, rabbit reticulocyte lysate (RLL), wheat germ, *E. coli*, and yeast lysate systems. Many embodiments of the current invention are also well-suited for use in hybrid systems where components of different systems are combined.

tRNAs aminoacylated on a 3' amide bond are reported not to combine with the elongation factor EF-TU which assists in binding to the A site (Sprinzl and Cramer, Prog. Nuc. Acid Res. 22:1 (1979), herein incorporated by reference). Such modified tRNAs do, however, bind to the A site. This binding of 3' modified tRNAs can be increased by changing the Mg<sup>++</sup> concentration (Chinali et al., Biochem. 13:3001 (1974), herein incorporated by reference). The appropriate concentrations and/or molar ratios of SATA and Mg<sup>++</sup> can be determined empirically. If the concentration or A site avidity of SATA is too high, the SATA could compete with native tRNAs for non-cognate codons i.e., could function much like puromycin and stall translation. If the concentration or A site avidity of SATA is too low, the SATA might not effectively compete with the release factors, i.e., it would not act as an effective nonsense suppressor tRNA. The balance between these can be determined empirically.

It is also believed that the elongation factor aids in proofreading the codon-anticodon recognition. The error rate in the absence of elongation factor and the associated GTP hydrolysis is estimated to be 1 in 100 for codons one nucleotide away

(Voet and Voet, Biochemistry 2nd ed. pp. 1000-1002 (1995), John Wiley and Sons, herein incorporated by reference). In a preferred embodiment, UAA is used as the linking codon. For UAA as the linking codon, there are 7 non stop codons which differ by one amino acid. This is 7/61 or about 11.5% of the non stop codons. One can estimate the probability of miscoding a given codon as  $(0.01)(0.115) = 1.15 \times 10^{-3}$  miscodes per codon. Thus, one would expect a miscode about every 870 codons, a frequency which will not substantially impair performance of various methods of the current invention. In an alternative embodiment, UAG or UGA is used as the linking codon.

In one embodiment, use of the mRNA with the selected codon at the end of the translatable reading frame would obviate this issue, e.g., by having it be the 3' end or having modifications to the more 3' moieties rendering them untranslatable and incapable of recognizing release factors, or by depleting the tRNAs cognate to any codons 3' of the linking codon. In an alternative embodiment, UAG or UGA is used as the linking codon.

In several embodiments, appropriate concentrations of SATA and  $Mg^{++}$  are used in the in vitro translation system, e.g. RRL, in the presence of the mRNA molecules in the pool, causing translation to cease when the ribosome reaches the codon which permits the SATA to accept the peptide chain (the linking codon described above). Within a short time, most of the linking codons will be occupied by SATAs within ribosomes. In a preferred embodiment, the system then will be irradiated with UV light, preferably at approximately 320 nm to 400 nm. Nucleic acids are typically transparent to, i.e. do not absorb, this wavelength range. Upon irradiation, the psoralen monoadduct will convert to a crosslink connecting the anticodon and the codon by a stable covalent bond.

In a preferred embodiment, the target mRNA is pre-selected. In another embodiment, the target mRNA is artificially produced. In an alternative embodiment, the target consists of messages native to the system under investigation, which may be unknown and/or unidentified. The ability to use unknown and/or unidentified mRNAs is a particular advantage of several embodiments of the current invention.

#### A Method for Producing Random or Quasi-random mRNA Libraries.

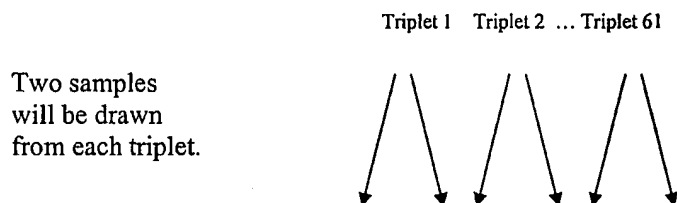
One difficulty with assembling messenger RNA's by random polymerization of nucleotides is that 3/64 or .047 of the codons that would occur randomly would be stop codons. Since the chance of not having a stop codon is 1-.047 or 0.953. This means that the chance of having a message N nucleotides in length would be  $(.953)^N$ . This can limit the length of messages of such production, as seen in Figure 7.

Thus the yield of messages of 100 nucleotide length is 0.008. The usual methods of producing these libraries is to produce cDNA's first and then transcribe them. In one embodiment of the current invention, RNA ligation is used. In one embodiment, RNA ligation using T4 RNA ligase is used. One advantage of using RNA ligation is the high yield, which may be reduced in some cases where secondary structure interferes.

In one embodiment, the method first acquires a library of codons, that is, it assembles highly pure triplets corresponding to the 61 sense codons but not including the three nonsense codons. These will be produced with an accuracy of 0.99 per nucleotide. Only 18 of the 61 codons can become stop codons by a single mutation. Of these, 5 have a 0.22 chance of becoming a stop codon in one mutation and 13 have a 0.11 chance of becoming a stop codon in one mutation. Or, 5/61 of the codons have a 0.22 chance of becoming a stop codon .01 of the time and 13/61 of the codons have a 0.11 chance of becoming a stop codon .01 of the time, giving  $5/61 \times 0.22 \times .01 = 1.80 \times 10^{-4}$  and  $13/61 \times 0.11 \times .01 = 2.34 \times 10^{-4}$  yielding a sum of  $4.15 \times 10^{-4}$  of mutating to a stop codon. To get 100 such codons attached would yield  $(4.15 \times 10^{-4})^{100} = 0.96$  yield. To protect against deletion or insertions, each triplet will be purified by anion exchange HPLC, an effective means of yielding high purity length discrimination. This should make the length purity at least 0.999. Again the yield for 100 codons would be  $(0.999)^{100} = 0.90$  yield.

In one embodiment, using these 61 different highly purified triplets the following procedure will be carried out:

Starting with NNNp triplets



One of these  
samples will  
become an "acceptor"  
and the other a "donor".

A	D	A	D	A	D
NNN Acceptor			pNNNp Donor		

In one embodiment, the acceptor will have no 5' or 3' phosphates and the donor will be treated to have both 5' and 3' phosphates. The acceptors will have the 3' phosphate removed as by T4 polynucleotide kinase with a buffer favoring its 3' phosphatase activity. The donors will have the 5' phosphate added by using the mutant T4 polynucleotide kinase lacking a 3'phosphatase activity and the appropriate buffer.

In one embodiment, all 61 acceptors and 61 donors will be combined to form the substrate for T4 RNA ligase. The proportion of each can be varied to change the bias of the resulting RNA constructs.

In one embodiment, under the action of T4 RNA ligase the 3' end of the acceptor will become attached to the 5' end of the donor. The result will be a 6 mer with a phosphorylated 3' end. One reason to have the 3' phosphorylation on the donor is to have no species that can be both a donor and an acceptor since this can lead to spontaneous circles of various sizes. The 6 mers can be purified by size again by using anion exchange HPLC. One of skill in the art will understand that the 3' phosphorylation can be located in other locations in accordance with several embodiments of the current invention.

These 6 mers will be divided into two samples, one dephosphorylated to form an acceptor and the other biphosphorylated to form a donor. The ligation step above is repeated and the resulting 12 mers purified by size. This can be repeated for a total of 4 or 5 cycles. At this point the anion exchange HPLC will lose its ability to discriminate by size and the length purification step will be omitted. The total number of steps, forming donors and acceptors and ligation, will go to 7 which will yield reading frames with an average of  $2^7$  or 128 codons. It is reasonable to expect 80% yields at each round yielding  $0.8^7$  or .2% yield. Since the triplets are commercially available in micromolar amounts, this will yield roughly  $10^{17}$  different random constructs. These will then be attached to 5' acceptor with a ribosome binding sequence and an AUG start codon. This construct will then be attached to a 3' donor with a stop codon cognate to our SATA or a linker consistent with the Phyllos or Nemoto methods,

usually containing a poly A tail. This yields an mRNA, or a construct for other RNA display technologies coding for a peptide with 128 amino acids. U.S. Patent No. 6,312,927 and 5,658,754, and the following references are herein incorporated by reference: (1) Basic Methods in Molecular Biology 2<sup>nd</sup> ed. Davis, Kuehl Battey; pub Appleton and Lange 1994; and (2) Romaniuk, P. J., Uhlenbeck O. C., Methods in Enzymology 100: 52-59 (1983), pub Appleton and Lange 1994.

In one embodiment of the present invention, the SATA has a puromycin on the 3' end and a crosslinker (such as psoralen) on the anticodon loop. In another embodiment, the SATA has a puromycin on the 3' end and the crosslinker is located on the mRNA. In some embodiments, where the crosslinker is on the mRNA, the crosslinker is positioned at a stop codon on the mRNA. In other embodiments, the crosslinker is located near a stop codon, preferably between about 1-20 nucleotides away, more preferably 1-10 nucleotides away, and most preferably 1-3 nucleotides away. One skilled in the art will understand that the crosslinker can also be designed to be placed more than 20 nucleotides away from the stop codon. As described herein, psoralen is one example of a crosslinker. Other crosslinkers are described herein.

In yet another embodiment, a Linking tRNA Analog is used to connect the mRNA to its cognate peptide. In one embodiment, the Linking tRNA Analog is a native or a synthetic tRNA (or a combination of native-synthetic hybrid) that has a crosslinker positioned on the anticodon loop. Preferably, the crosslinker is bound to the anticodon loop through covalent bonding. In one embodiment, the Linking tRNA Analog accepts the nascent peptide onto its 3' aminoacyl moiety through the action of ribosomal peptidyl transferase. The 3' aminoacyl moiety can be native to the tRNA or can be synthetically introduced. In one embodiment, the ester bond between the peptide and the tRNA is protected from ribosomal peptidyl transferase because the message is untranslatable beyond the codon bound by the tRNA (the linking codon). Thus, the ribosomal peptidyl transferase will be unable to release the peptide from the tRNA. Therefore, in several embodiments of the present invention, the ester bond between the tRNA and a peptide chain is rugged enough to obviate the need for puromycin. The connection between the Linking tRNA Analog and the peptide, when linked through an ester bond, is protected from dissolution by ribosomal peptidyl transferase by making the translated message "untranslatable" beyond the linking

codon. Advantageously, the message then will be stably attached to its peptide for further identification, selection and evolution. Another advantage is that synthetic or modified tRNAs need not be used in some embodiments employing the Linking tRNA Analog. In one particular embodiment, the tRNA is unmodified in the sense that it is unmodified on the 3' end, and may or may not have minor modifications on the anticodon loop. In many embodiments, unmodified native tRNA (particularly unmodified on the 3' end) can be used, therefore making the system, among other things, more cost-effective, efficient, quicker, less error-prone, and capable of producing a much higher yield. Not wishing to be bound by the following theory, the inventors believe that absence of puromycin (or similar linkers) results, in some cases, in low yield because puromycin obstructs the interaction of the elongation factor with tRNA thus affecting yield. Further, the elongation factor, when unobstructed by puromycin (or similar linkers) is able to accomplish dynamic proof-reading, thereby reducing error rates.

In a further embodiment, a Nonsense Suppressor tRNA is used. The Nonsense Suppressor tRNA recognizes a stop codon or a pseudo stop codon. The Nonsense Suppressor tRNA is used to connect the mRNA to its cognate peptide. In one embodiment, the Nonsense Suppressor tRNA is a native or a synthetic tRNA (or a combination of native-synthetic hybrid). In one embodiment, the Nonsense Suppressor tRNA has an anticodon triplet that hydrogen bonds to a stop or pseudo stop codon. In one embodiment, the Nonsense Suppressor tRNA has 3' modifications and sequences that conform to the Yarus extended anticodon rules (Yarus, Science 218:646-652, 1982, herein incorporated by reference). In one embodiment, the Nonsense Suppressor tRNA Analog accepts the nascent peptide onto its 3' aminoacyl moiety through the action of ribosomal peptidyl transferase. The 3' aminoacyl moiety can be native to the tRNA or can be synthetically introduced. In one embodiment, the ester bond between the peptide and the tRNA is protected from ribosomal peptidyl transferase because the message is untranslatable beyond the codon bound by the tRNA (the linking codon). Thus, the ribosomal peptidyl transferase will be unable to release the peptide from the tRNA. In a preferred embodiment, the Nonsense Suppressor tRNA does not have any type of crosslinker: the crosslinker is instead located on the mRNA. In some embodiments, where the crosslinker is on the mRNA, the crosslinker

is positioned at or near a stop codon on the mRNA. Therefore, several embodiments of the present invention offer several advantages. For example, the surprisingly rugged ester bond between the Nonsense Suppressor tRNA and the means that a puromycin, a puromycin analog, or other amide linker is not needed. Another advantage is that the linkage between the Nonsense Suppressor tRNA and the peptide, when linked through an ester bond, is protected from dissolution by ribosomal peptidyl transferase by making the translated message "untranslatable" beyond the linking codon. Advantageously, the message then will be stably attached to its peptide for further identification, selection and evolution. Thus, in several embodiments, the Nonsense Suppressor tRNA does not need a puromycin nor a crosslinker positioned on the tRNA itself. Yet another advantage is that synthetic or modified tRNAs need not be used. In one particular embodiment, the tRNA is unmodified in the sense that it is unmodified on the 3' end, and may or may not have minor modifications on the anticodon loop. In many embodiments, unmodified native tRNA (particularly unmodified on the 3' end) can be used, therefore making the system, among other things, more cost-effective, efficient, quicker, less error-prone, and able to offer a high yield. Not wishing to be bound by the following theory, the inventors believe that absence of puromycin (or similar linkers) results, in some cases, in low yield because puromycin obstructs the interaction of the elongation factor with tRNA thus affecting yield. Further, the elongation factor, when unobstructed by puromycin (or similar linkers) is able to accomplish dynamic proof-reading, thereby reducing error rates.

A preferred embodiment of the invention comprises a tRNA molecule capable of covalently linking a nucleic acid encoding a polypeptide and the polypeptide to the tRNA. In one embodiment, the linkage of the nucleic acid occurs on a portion of the tRNA other than the linkage to the polypeptide and the tRNA comprises a linking molecule associated with the anticodon of the tRNA. This anticodon of the tRNA is capable of forming a crosslink to the mRNA under irradiation with light of a required wavelength, preferably a furan-sided psoralen monoadduct on the anticodon irradiated with UVA, preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. In one embodiment, an amino acid or amino acid analog is attached to the 3' end of a tRNA molecule by a stable bond to generate a SATA. One advantage of some embodiments of the invention

is that it ensures that the translation process stalls at this point, thereby making the bond stable in subsequent applications.

In one embodiment, the anticodon of the tRNA is capable of forming a crosslink to the mRNA, where the cross-link is a non-psoralen crosslinker molecule or moiety. As used herein, the term "non-psoralen crosslinker" shall be given its ordinary meaning and shall include one or more of the following compounds: 2-thiocytosine, 2-thiouridine, 4-thiouridine, 5-iodocytosine, 5-iodouridine, 5-bromouridine, 2-chloroadenosine, aryl azides, and modifications or analogues thereof.

Other embodiments include an mRNA comprising a psoralen, or a non-psoralen crosslinker, preferably located in the 3' region of the reading frame, more preferably at the most 3' codon of the reading frame, most preferably at the 3' stop codon of the reading frame. In preferred embodiments, the linkage between the tRNA and the mRNA is a cross-linked psoralen, or a non-psoralen crosslinker molecule. In one embodiment, the linkage between the tRNA and the mRNA is a furan-sided psoralen monoadduct.

In several embodiments, the present invention permits the attachment of a protein to its respective mRNA without requiring any or substantial modification of native tRNA. In one embodiment, modified tRNA is used.

One embodiment of the invention comprises an mRNA molecule capable of covalently linking a tRNA that is covalently linked to a polypeptide encoded by the mRNA wherein the tRNA comprises a linking molecule associated with the codon of the mRNA. This codon of the mRNA is capable of forming a crosslink to the tRNA under irradiation with light of a required wavelength. The moiety, which is driven to crosslink, is preferably a furan-sided psoralen monoadduct, or a non-psoralen crosslinker on the codon irradiated with UVA, preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. Preferably, this codon is the last (3' most) translatable codon of the reading frame and hence stops translation and is a stop or pseudo stop codon. By making the mRNA untranslatable beyond this point, the use of a bond between the tRNA or tRNA analog and the encoded peptide that is stable to the peptidyl transferase is unnecessary to stall the translation. For many applications, the native ester bond is adequately stable. In one embodiment, the message is made untranslatable by one or more of the



following techniques: (1) making the codon the physical end; (2) by using modified nucleotides; (3) by using moieties that can not be processed by the ribosome; and (4) by depleting the tRNAs recognizing the message beyond the selected codon. One of skill in the art will understand that other methods that render the message untranslatable can also be used in accordance with several embodiments of the invention.

One skilled in the art will understand that, in accordance with some embodiments of the present invention, other methods to crosslink an mRNA to a translating tRNA while still in the ribosome can also be used. These methods include, but are not limited to, the use of modified nucleotides such as aryl azides on uracils and guanine residues which provide efficient mRNA-tRNA photo crosslinks in ribosomes (Demeshkina, N, *et al.*, RNA 6:1727-1736, 2000, herein incorporated by reference).

A further embodiment of the invention provides a method of forming a monoadduct. According to one embodiment, a target oligonucleotide with at least one uridine and at least one modified uridine is contacted with psoralen, and the target oligonucleotide and psoralen are coupled to form a monoadduct. The modified uridine according to this embodiment may be modified to avoid coupling with psoralen. In one embodiment, the modified uridine is pseudouridine. According to this embodiment, the target oligonucleotide may be a tRNA molecule, such as tRNA, modified tRNA and tRNA analogs or a mRNA molecule, such as mRNA, modified mRNA and mRNA analogs. In a further embodiment the psoralen is coupled to the target oligonucleotide by one or more cross-links. According to this embodiment, a second oligonucleotide with a nucleotide sequence complementary to the target oligonucleotide sequence may be present. This second oligonucleotide may contain no uridine or may contain uridine residues that are modified to avoid cross-linking with the target oligonucleotide. Preferably, the modified uridine is pseudouridine.

In one embodiment of the present invention, the invention comprises a method of forming a psoralen monoadduct on a nucleic acid. The method, in some embodiments, comprises providing a first nucleic acid and a second nucleic acid that are at least substantially complementary to each other. The first nucleic acid comprises one or more uridine monoadduct targets, and the second nucleic acid comprises at least one pseudouridine. The method further comprises hybridizing at least a portion of the first nucleic acid and the second nucleic acid in the presence of psoralen, or psoralen-

like agent, to form a hybrid, irradiating the hybrid with ultraviolet light, thereby forming the psoralen monoadduct on the first nucleic acid. In one embodiment, one or more uridine monoadduct targets comprises a uridine located adjacent to an adenosine, preferably 3' from the adenosine.

In another embodiment, a method of producing a psoralen monoadduct or a crosslink comprises providing a first nucleic acid and a second nucleic acid that are at least substantially complementary to each other. The first nucleic acid comprises one or more uridine monoadduct targets or crosslink targets and one or more uridine monoadduct non-targets or crosslink non-targets. The uridine monoadduct non-targets or crosslink non-targets are operable to be replaced or substituted with one or more pseudouridines. The method further comprises replacing one or more of the uridine monoadduct non-targets or crosslink non-targets with pseudouridine, hybridizing at least a portion of the first and second nucleic acids in the presence of psoralen, forming at least a partial hybrid; and irradiating, or otherwise activating, the hybrid, thereby forming the psoralen monoadduct or the crosslink on the first nucleic acid on the targets, while protecting the nontargets. In one embodiment, visible light is used to form the adduct or crosslink. In another embodiment, ultraviolet light is used.

Several embodiments of the present invention include a method of stably linking a nucleic acid, a tRNA, and a polypeptide encoded by the nucleic acid together to form a linked nucleotide-polypeptide complex. In a preferred embodiment, the nucleic acid is an mRNA and the linked nucleotide-polypeptide complex is a mRNA-polypeptide complex. The method can further comprise providing a plurality of distinct nucleic acid-polypeptide complexes, on, for example, an array, providing a ligand with a desired binding characteristic, contacting the complexes with the ligand, removing unbound complexes, and recovering complexes bound to the ligand.

Several methods of the current invention involve the identification, selection and/or evolution of nucleic acid molecules and/or proteins. In one embodiment, this invention comprises amplifying the nucleic acid component of the recovered complexes and introducing variation to the sequence of the nucleic acids. In other embodiments, the method further comprises translating polypeptides from the amplified and varied nucleic acids, linking them together using tRNA, and contacting them with the ligand to select another new population of bound complexes. Several embodiments of the

present invention use selected protein-mRNA complexes in a process of *in vitro* evolution, in particular the iterative process in which the selected mRNA is reproduced with variation, translated and again connected to cognate protein for selection.

In one embodiment of the current invention, the selected codon is located on, or placed at, the end of the translatable reading frame by having it be the 3' end or having modifications to the more 3' moieties rendering them untranslatable and incapable of recognizing release factors. One advantage of this embodiment is that the amide bonded amino acid analog on the 3' end of the tRNA is not needed to stall translation. Further, this permits efficient production of peptide tRNA complexes. These complexes are quite robust in spite of the high energy content of their ester bond (Figure 6).

In a preferred method, the SATA or peptidyl-tRNA will be attached to the translated message by a psoralen, or one of the group 2-thio cytosine, 2-thio uridine, 4-thio uridine 5-iodocytosine, 5-iodouridine, 5-bromouridine, 2-chloroadenosine, or aryl azides cross link between the codon and anticodon. Psoralen cross links are, in some embodiments, preferentially made between sequences that contain complementary 5' pyrimidine-purine 3' sequences, especially UA or TA sequences (Cimino *et al.*, Ann. Rev. Biochem. 54:1151 (1985), herein incorporated by reference). In some embodiments, non-psoralen crosslinkers or aryl azides are used and in certain embodiments, are particularly advantageous because they are less stringent in their requirements and therefore increase the possible codon-anticodon pairs.

The codon coding for the SATA or the Linking tRNA Analog may be referred to as the linking codon. For the use of psoralen as the crosslinking moiety, the linking codon can be PYR-PUR-X or X-PYR-PUR, so that several codons may be used for the linking codon. "X" in this case, may be any nucleotide. Conveniently, the stop or nonsense codons have this configuration. Using a codon that codes for an amino acid may require minor adjustments to the genetic code, which could complicate some applications. Therefore, in a preferred embodiment, a stop codon is used as the linking codon and the SATA or linking tRNA functions as a nonsense suppressor in that it recognizes the linking codon. One skilled in the art, however, will appreciate that, with appropriate adjustments to the system, any codon can be used.

In several embodiments, once all the nascent proteins are connected to their cognate mRNAs, the ribosomes are released or denatured. Preferably, this is

accomplished by the depletion of  $Mg^{++}$  through dialysis, simple dilution, or chelation. One skilled in the art will understand that other methods, including, but not limited to, denaturation by changing the ionic strength, the pH, or the solvent system can also be used.

In several embodiments of the invention, the selection of cognate pairs will be based upon affinity binding of proteins according to any of a variety of established methods, including, but not limited to, arrays, affinity columns, immunoprecipitation, and many high throughput screening procedures. A variety of ligands may also be used, including, but not limited to, proteins, nucleic acids, chemical compounds, polymers and metals. In addition, cell membranes or receptors, or even entire cells may be used to bind the cognate pairs. The selection can be positive or negative. That is, the selected cognate pairs can be those that do bind well to a ligand or those that do not. For instance, for a protein to accelerate a thermodynamically favorable reaction, e.g., act as an enzyme for that reaction, it should bind both the substrate and a transition state analog. However, the transition state analog should be bound much more tightly than the substrate. This is described by the equation

$$\frac{k_{enzyme}}{k_{enzyme}} = \frac{K_{trans}}{K_{subst}}$$

where the ratio of the rate of the reaction with the enzyme,  $k_{enzyme}$ , to the rate without,  $k_{enzyme}$ , is equal to the ratio of the binding of the transition state to the enzyme  $K_{trans}$  over the binding of the substrate to the enzyme  $K_{subst}$  (Voet and Voet, Biochemistry 2nd ed. p.380, (1995), John Wiley and Sons, herein incorporated by reference).

In a preferred embodiment, proteins which compete poorly for binding to the substrate but compete well for binding to the transition state analog are selected. Operationally, this may be accomplished by taking the proteins that are easily eluted from a matrix with substrate or substrate analog bound to it and are the most difficult to remove from matrix with transition state analog bound to it. By sequentially repeating this selection and reproducing the proteins through replication and translation of the nucleic acid of the cognate pairs, an improved enzyme should evolve. Affinity to one

entity and lack of affinity to another in the same selection process is used in several embodiments of the current invention. Selection can also be done by RNA in many embodiments.

Once the selection has identified a population of cognate pairs it may be convenient to detach the mRNA strand from the tRNA molecule to reproduce it. This is not always necessary, but when desired in certain embodiments, can be accomplished by using psoralen as the connecting photolinker and irradiating the pairs with UV, preferably at approximately 313 nm or just below. This has been identified as a wave length that will photoreverse the psoralen crosslink to MAf and damage the nucleic acid minimally. The ratio of photoreversal to nucleic acid damage is estimated to be 1 photoreversal for damage to 1 in 600 bases (Cimino *et al.*, Biochem 25:3013 (1986), herein incorporated by reference).

One skilled in the art will appreciate that the mRNAs can be reproduced in many ways including, but not limited to, by RNA-dependent RNA polymerases or by reverse transcription and PCR. This can take place using mRNAs separated from the cognate pairs, e.g., using poly T or poly U to hybridize to the poly A tails of, for instance, native unknown messages or by leaving the cognate pairs intact and using oligonucleotide primers that hybridize partially into the reading frame for known messages. Alternatively, commercial kits for rapid amplification of cDNA ends may be used. In several embodiments, the methods described above for placement of photoactivatable moieties on oligonucleotides can be used to create modified oligoribonucleotides which can then be attached to the 3' ends of the message using T4 RNA ligase. The oligonucleotides attached would contain the linking codon with its photoactivatable moiety.

As described herein, there are several ways to connect the message to the tRNA in accordance with several embodiments of the present invention. For example, the following table outlines some embodiments of the current invention:

Stable Acceptor

Native Esterified Acceptor

Crosslinker on tRNA Analog	<i>tRNA Analog Characteristics:</i> 1) Stable acceptor 2) Anticodon loop crosslinker 3) Recognizes linking codon  <i>mRNA Characteristics:</i> 1) Flexible; can be a stop or a pseudo stop codon	<i>tRNA Analog Characteristics:</i> 1) aaRS for aminoacylating or chemical aminoacylation 2) Anticodon loop crosslinker 3) Recognizes linking codon  <i>mRNA Characteristics:</i> 1) Untranslatable beyond linking codon.
Crosslinker on mRNA	<i>tRNA Analog Characteristics:</i> 1) Stable acceptor 2) Recognizes linking codon   <i>mRNA Characteristics:</i> 1) Crosslinker on or near linking codon	<i>tRNA Analog Characteristics :</i> 1) Recognizes linking codon 2) Means to aminoacylate (Native nonsense suppressor can work)  <i>mRNA Characteristics:</i> 1) Contains linking codon 2) Untranslatable beyond linking codon 3) Crosslinker on or near linking codon

In one embodiment, at least one amino acid substitution at each position in the protein is sampled. This is particularly advantageous for the evolution of proteins.

#### **The Replication Threshold**

A nominal minimum number of replications for efficient evolution may be estimated using the following formulae. If there is a sequence which is  $n$  sequences in length, with a selective improvement  $r$  mutations away with a mutation rate of  $p$ , the probability of generating the selective improvement on replication may be determined as follows:

For  $r=1$ , probability of a mutation at the right point,  $p$ , times the probability that it mutated to the right one of the three nucleotides that are different from the starting point,  $1/3$ , times the probability that the other  $n-1$  sites remain unmutated,  $(1-p)^{(n-1)}$ , or

$$P_r = \left(\frac{p}{3}\right)^1 (1-p)^{(n-1)}$$

where,  $P$  = the probability of attaining a given change  $r$  mutations away. More generally, for all  $r$  values:

$$P_r = \left(\frac{p}{3}\right)^r (1-p)^{(n-r)}$$

It is instructive to compare the chances of finding an advantage one mutation away with the chances three mutations away. This is because, given the triplet genetic code, any given codon can only change into nine other codons in one mutation. Indeed, it turns out that no codon can actually change into nine other amino acid codes in one mutation. The maximum number of amino acids that can be accessed in one mutation is seven amino acids and there are only eight codons of the sixty-four that can do this. Most codons have five or six out of nineteen other amino acids within one mutation. To reach all nineteen amino acids that are different from the starting one requires, in general, three mutations. These three mutations cannot be sequential since the two intervening ones will not, in general, be selectively advantageous. Therefore we need to use steps that are, at least, three mutations in size ( $r=3$ ) to use all 20 amino acids.

For a mutation rate of .0067, which is that reported for “error-prone PCR”, using a message of 300 nucleotides, which gives a short protein of 100 amino acids:

$$B = 1.51 \times 10^{-9}$$

Therefore, one would expect to need a threshold of:

$$\frac{1}{1.51 \times 10^{-9}} = 6.64 \times 10^8$$

replications at that mutation rate to reasonably expect to reach the next amino acid that is advantageous. This is not the replication to use since the binomial expansion shows that over 1/3 of trials (actually about 1/e) would not contain the given sequence with selective advantage.

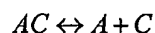
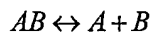
A poisson approximation for large  $n$  and small  $p$  for a given  $\mu$  can be calculated so that we can compute the general term when  $n$  is, say, of the order  $10^9$  and  $p$  is of the order  $10^{-9}$ . The general term of the approximation is:

$$\frac{\mu^r}{r!e^\mu}$$

An amplification factor of greater than approximately 6/P ensures that evolution will progress with the use of all amino acids. This is useful when the production of novel proteins precludes the use of “shuffling” of preexisting proteins.

### Limits on Purification

Given a reversible binding where B and C compete for A:



$$k_c = \frac{[A][C]}{[AC]}$$

$$[B] = k_b \frac{[AB]}{[A]} \quad (1)$$

$$[C] = k_c \frac{[AC]}{[A]} \quad (2)$$

The total concentrations can be expressed as follows:

$$[B]_r = [B] + [AB] \quad (3)$$

$$[C]_r = [C] + [AC] \quad (4)$$

Dividing (3) by (4):

$$\frac{[B]_r}{[C]_r} = \frac{[B] + [AB]}{[C] + [AC]} \text{ And substituting (1) and (2) for [B] and [C]:}$$

$$\frac{[B]_r = k_b \left[ \frac{AB}{A} \right] + [AB]}{[C]_r = k_c \left[ \frac{AC}{A} \right] + [AC]} :$$

Rearranging the equation gives the following results:



$$\frac{[B]_T}{[C]_T} = \frac{[AB] \left( \frac{k_B + [A]}{[A]} \right)}{[AC] \left( \frac{k_C + [A]}{[A]} \right)}$$

Canceling the [A]'s in the numerator and denominator the equation gives the following results:

$$\frac{[B]_T}{[C]_T} = \frac{[AB](k_B + [A])}{[AC](k_C + [A])}$$

Finally, rearranging the equation provides the following equation:

$$\frac{[AB]}{[AC]} = \frac{[B]_T(k_C + [A])}{[C]_T(k_B + [A])}$$

$$\frac{(k_C + [A])}{(k_B + [A])} \quad (\text{Enrichment Factor})$$

The above factor is termed the "Enrichment Factor". The ratio of the total components is multiplied by this factor to calculate the ratio of the bound components, or the enrichment of B over C. The maximum enrichment factor is  $k_C/k_B$ , when the [A] is significantly smaller than  $k_C$  or  $k_B$ . When [A] is significantly greater than  $k_C$  or  $k_B$ , the enrichment is 1, that is, there is no enrichment of one over the other.

The enrichment is limited by the ratio of binding constants. To enrich a scarce protein that is bound 100 times as strongly as its competitors, the ratio of that protein to its competitors is increased by 1 million with 3 enrichments. To enrich a protein that only binds twice as strongly as its competitors, 10 enrichment cycles would gain only an enrichment of ~1000.

By an exactly analogous method an enrichment factor of selecting proteins that bind least well can be shown:

In the equation:

$$\frac{[C]}{[B]} = \frac{k_C[C]_T([A] + k_B)}{k_B[B]_T([A] + k_C)}$$

The enrichment here is maximal at  $[A] > k_A$  or  $k_B$ .

$$\frac{k_C ([A] + k_B)}{k_B ([A] + k_C)}$$

### Strategy For Selection

When dealing with populations of molecules, the selection criterion is often affinity or strength of binding to a target ligand. What are the limits of this selection? Consider a population of molecules,  $A_i$ , (the  $i$ 's correspond to different binding affinities) that will compete for binding to a target ligand,  $T$ .

The reactions can be represented:  $A_i T \rightleftharpoons A_i + T$ . Dissociation constant ( $k_d$ ) values are derived by the equation below:

$$k_i = \frac{[A_i][T]}{[TA_i]}$$

Dissociation constant is used instead of binding constant because it conveniently has the same units as the concentration.

If one expresses the total amount of  $A_i$  present as the sum of the bound and the unbound  $A_i$ , the equation can be expressed as follows:

$$[A_i]_{tot} = [A_i] + [A_i T]$$

$$[A_i] = [A_i]_{tot} - [A_i T]$$

Substituting and rearranging, we can express the fraction of the total of  $A_i$  that is bound as follows:

$$\frac{[A_i T]}{[A_i]_{tot}} = \frac{[T]}{(k_i + [T])}$$

Notice that when the concentration of free or unbound T ( $[T]$ ) is large compared to the dissociation constant, the fraction of bound goes to one (see Figures 9 & 10). That is, all of the  $A_i$  with that  $k$  value is bound. For example, if all of the binding proteins in a population are of equal concentration for all  $K$  values between 1 and  $10^{12}$ , the data generates a flat curve as shown in Figure 10.

However, if the tighter binding proteins are enriched by letting them come to equilibrium with a target ligand until an unbound target concentration,  $[T]$ , of  $10^{-6}$  is reached, the distribution of bound proteins would generate a sigmoid curve as shown above.

Since the recovery by binding depends on the unbound T at equilibrium, it is convenient to be able to know how much total target to add to yield a given unbound or free T. The total T is equal to the sum of the bound and the unbound T as shown in the following equation:

$$[T]_{total} = [T] + \sum [TA_i]$$

The  $[T]$  is the free T that we select and the  $\sum [TA_i]$  is the amount under the curve computed by knowing the original distribution as above. In many cases the area under the curve is small compared to the selected  $[T]$  and the total T can be closely approximated by the free or unbound T.

Now consider a more realistic distribution. If one produces proteins that are more or less random in sequence as by translating mRNA's that are more or less random in sequence, and then exposes the resulting proteins to a ligand, what would the

affinity distribution of the population of proteins to that ligand look like? In other words, if one plots the concentration vs. the association constant, what sort of curve would result? One would expect it to follow the description of Burnet (16) in 1963, "...most of the molecules will show a minimal adsorptive affinity... while only an occasional combining site pattern will show a high affinity." This has been refined by Lancet et al (17), who suggest that this distribution could be expected to be a binomial of the logarithm of the association constant. They derived values for the parameters of such a distribution empirically, using a population of immunoglobulins that were naive to the ligand. They showed that, in many ranges of the parameters, this binomial approaches a Poisson distribution as shown in Figure 11.

One can use this distribution to look at the family of curves that would be bound by different [T] values as demonstrated in Figure 12.

Sufficiently stringent or high p[T] values will yield areas under the bound curve that will be significantly less than [T] so that [T] can be used as an approximation of [T]<sub>tot</sub>.

Lastly, one must consider the necessary stringency. If a mixture containing two populations of proteins A<sub>i</sub> and A<sub>j</sub> with dissociation constants k<sub>i</sub> and k<sub>j</sub>, ( k<sub>i</sub> < k<sub>j</sub> ) are considered, how well can they be separated? This can be determined by applying the equation 1 above to both proteins as shown in the following equations:

$$[A_i T] = \frac{[A_i]_{tot} [T]}{(k_i + [T])} \quad [A_j T] = \frac{[A_j]_{tot} [T]}{(k_j + [T])}$$

Taking the ratio of the two expressions:

$$\frac{[A_i T]}{[A_j T]} = \frac{\frac{[A_i]_{tot} [T]}{(k_i + [T])}}{\frac{[A_j]_{tot} [T]}{(k_j + [T])}}$$

Simplifying,

$$\frac{[A_i T]}{[A_j T]} = \frac{[A_i]_{tot}}{[A_j]_{tot}} \frac{(k_j + [T])}{(k_i + [T])}$$

↑ Ratio after enrichment by binding
 ↑ Ratio before enrichment
 ↑ "Enrichment Factor"

This gives the ratio of the two species after enrichment in terms of the starting ratio and a factor called the “enrichment factor”:

$$\frac{(k_j + [T])}{(k_i + [T])}$$

Notice that the enrichment factor is maximal when [T] is small compared to the k values and its maximum is  $k_j/k_i$ , but when [T] is large compared to the k values there is no enrichment and the factor is 1. This means that the ability to enrich cannot be greater than the ratio of the k values.

Now, consider a real example of separation or purification by affinity binding. Xu et al (18) took 14 generations to isolate a tightly binding protein for tumor necrosis factor- $\alpha$ . Their protein had a k value of ~100 pM or  $10^{-10}$  M. If the original distribution is represented by a Poisson distribution, as per Lancet (17), the question becomes whether the selection stringency could have been more efficient. The answer is yes. By using a [T] value smaller than the goal k value, one can accomplish high purification of tightly binding proteins in four rather than 14 selection rounds as shown in Figure 13.

For the comparison above we reconstructed the original distribution from their published data (not shown). This example gives a method for more rapidly isolating a small number of high affinity proteins. However, in many applications it would be advantageous to evolve large numbers of proteins that have high binding. In those cases, less stringent [T] values would be used in combination with mutation steps. For example, rather than select for proteins with the highest binding constants, one can de-select for a population of proteins with low binding constants, e. g. proteins with binding constants between the red lines as shown.

Rather than select for a few proteins with the highest binding affinities in a given distribution one can use a less stringent selection so as to have a high number of

different sequences and use multiple rounds of mutation with gradual increase in the stringency to evolve a large population of proteins with a high binding affinity.

The following Examples illustrate various embodiments of the present invention and are not intended in any way to limit the invention.

#### **EXAMPLE 1: PRODUCTION OF THE SATA USING URIDINE**

One skilled in the art will understand that the SATA can be produced in a number of different ways. The protocols described below in the following examples can be used for SATAs that have both a puromycin and a crosslinker on the tRNA, or that have a puromycin on the tRNA and a crosslinker on the mRNA. Where the crosslinker is on the mRNA, Example 4, below, provides guidance. The following protocol is also instructive for Linking tRNA Analogs, in the sense that Linking tRNA Analogs also, in a preferred embodiments, have a crosslinker on the tRNA.

For example, in a preferred embodiment, three fragments (Fig. 1) were purchased from a commercial source (e.g., Dharmacon Research Inc., Boulder, CO). Modified bases and a fragment 3 with a pre-attached puromycin on its 3' end and a PO<sub>4</sub> on its 3' end were included, all of which were available commercially. Three fragments were used to facilitate manipulation of the fragment 2 in forming the monoadduct.

Yeast tRNA<sup>Ala</sup> or yeast tRNA<sup>Phe</sup> were used; however, sequences can be chosen from widely known tRNAs or by selecting sequences that will form into a tRNA-like structure. Preferably, sequences with only a limited number of U's in the portion that corresponds to the fragment 2 are used. Using a sequence with only a few U's is not necessary because psoralen preferentially binds 5'UA3' sequences (Thompson J.F., et al Biochemistry 21:1363, herein incorporated by reference). However, there would be less doubly adducted product to purify out if such a sequence was used.

Fragment 2 was preferably used in a helical conformation to induce the psoralen to intercalate. Accordingly, a complementary strand was required. RNA or DNA was used, and a sequence, such as poly C to one or both ends, was added to facilitate separation and removal after monoadduct formation was accomplished.

Fragment 2 and the cRNA were combined in buffered 50 mM NaCl solution. The T<sub>m</sub> was measured by hyperchromicity changes. The two molecules were re-

annealed and incubated for 1 hour with the selected psoralen at a temperature  $\sim 10^{\circ}\text{C}$  less than the  $T_m$ . The psoralen was selected based upon the sequence used. A relatively insoluble psoralen, such as 8 MOP, could be selected which has a higher sequence stringency but may need to be replenished. A more soluble psoralen, such as AMT, has less stringency but will fill most sites. Preferably, HMT is used. If a fragment 2 is chosen that contains more non-target U's, a greater stringency is desired. Decreasing the temperature or increasing ionic strength by adding  $\text{Mg}^{++}$  was also used to increase the stringency. In a preferred embodiment,  $\text{Mg}^{++}$  was omitted and  $\sim 400$  mM NaCl solution was used.

Following incubation, psoralen was irradiated at a wavelength greater than approximately 400 nm. The irradiation depends on the wavelength chosen and the psoralen used. For instance, approximately 419 nm 20-150 J/cm<sup>2</sup> was preferably used for HMT. This process results in an almost entirely furan sided monoadduct.

#### **PURIFICATION OF A MONOADDUCT**

The monoadduct was then purified by HPLC as described in Sastry et al, J. Photochem. Photobiol. B Biol. 14:65-79, herein incorporated by reference. The fact that fragment 2 was separate from fragment 3 facilitated the purification step because, generally, purification of monoadducts  $\geq 25$  mer is difficult (Spielmann *et al.* PNAS 89: 4514-4518, herein incorporated by reference).

#### **LIGATION OF FRAGMENT 2 AND 3**

The fragment 2 was ligated to the fragment 3 using T4 RNA ligase. The puromycin on the 3' end acted as a protecting group. This is done as per Romaniuk and Uhlenbeck, Methods in Enzymology 100:52-59 (1983), herein incorporated by reference. Joining of fragment 2+3 to the 3' end of fragment 1 was done according to the methods described in Uhlenbeck, Biochemistry 24:2705-2712 (1985), herein incorporated by reference. Fragment 2+3 was 5' phosphorylated by polynucleotide kinase and the two half molecules were annealed.

In an alternative method, significant quantities of furan sided monoadducted U were formed by hybridizing poly UA to itself and irradiating as above. The poly UA was then enzymatically digested to yield furan sided U which was protected and incorporated into a tRNA analog by nucleoside phosphoramidite methods. Other methods of forming the psoralen monoadducts include the methods described in

Gamper *et al.*, J. Mol. Biol. 197: 349 (1987); Gamper *et al.*, Photochem. Photobiol. 40:29, 1984; Sastry *et al.*, J. Photochem. Photobiol. B Biol. 14:65-79; Spielmann *et al.* PNAS 89:4514-4518, U.S. Patent Number 4,599,303, all herein incorporated by reference.

SATAs generated by the methods described above read UAG (anticodon CUA). Additionally, UAA or UGA were also used. In various embodiments, any message that had the stop codon that was selected as the "linking codon" was used.

#### **Production Of Psoralenated Furan Sided Monoadducts**

##### ***UV Light Exposure Of RNA:DNA Hybrids***

Equal volumes of 3 ng/ml RNA:cRNA hybrid segments and of 10 µg/ml HMT both comprised of 50mM NaCl were transferred into a new 1.5 ml capped polypropylene microcentrifuge tube and incubated at 37°C for 30 minutes in the dark. This was then transferred onto a new clean culture dish. This was positioned in a photochemical reactor (419 nm peak Southern New England Ultraviolet Co.) at a distance of about 12.5 cm so that irradiance was ~6.5 mW/cm<sup>2</sup> and irradiated for 60-120 minutes.

##### ***Removal of Low Molecular Weight Protoproducts***

100µl of chloroform-isoamyl alcohol (24:1) was pipetted and mixed by vortex. The mixture was centrifuged for 5 minutes at 15000 xg in a microcentrifuge tube. The chloroform-isoamyl alcohol layer was removed with a micropipette. The chloroform-isoamyl alcohol extraction was repeated once again. Clean RNA was precipitated out of the solution.

##### ***Alcohol Precipitation***

Two volumes (~1000 µl) ice cold absolute ethanol was added to the mixture. The tube was centrifuged for 15 minutes at 15,000xg in a microcentrifuge. The supernatant was decanted and discarded and the precipitated RNA was redissolved in 100µl DEPC treated water then re-exposed to the RNA+8-MOP.

##### ***Isolation of the Psoralenated RNA Fragments Using HPLC***

All components, glassware and reagents were prepared so that they were RNAase free. The HPLC was set up with a Dionex DNA PA-100 package column. The psoralenated RNA:DNA hybrid was warmed to 4°C. The psoralenated RNA was applied to HPLC followed by oligonucleotide analysis, as described in the following



section entitled "Oligonucleotide Analysis by HPLC." The collected fractions represented:

- 5'CUAGAΨCUGGAGG3'5'CUAGAΨCUGGAGG3', where Ψ is pseudouridine (SEQ ID NO: 1)
- Furan sided 5'CUPsoraleAGAGAΨCUGGAGG3' monoadducts (SEQ ID NO: 2)
- 5'XXXXXXCCUCCAGAUCUAGXXXXX3' (SEQ ID NO: 3)
- 5'XXXXXXCCUCCAGAUCUPsoraleAGXXXXX3' (SEQ ID NO: 4)

The fractions were stored at 4°C in new, RNAase free snapped microcentrifuge tubes and stored at -20°C if more than four weeks of storage were required.

***Identification of the RNA Fragments Represented by Each Peak Fraction Collected by HPLC Using Polyacrylamide Gel Electrophoresis (PAGE)***

The electrophoresis unit was set up in a 4°C refrigerator. A gel was selected with a 2 mm spacer. Each 5 µl of HPLC fraction was diluted to 10 µl with Loading Buffer. 10 µl of each diluted fraction was loaded into appropriately labeled sample wells. The tracking dye was loaded in a separate lane and electrophoresis was run as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." After the electrophoresis run was complete, the electrophoresis was stopped when the tracking dye reached the edge of the gel. The apparatus was disassembled. The gel-glass panel unit was placed on the UV light box. UV lights were turned on. The RNA bands were identified. The bands appeared as denser shadows under UV lighting conditions.

***Extraction of the RNA From the Gel***

Each band was excised with a new sterile and RNAase free scalpel blade and transferred into a new 1.5 ml snap capped microcentrifuge tube. Each gel was crushed against the walls of the microcentrifuge tubes with the side of the scalpel blade. A new blade was used for each sample. 1.0 ml of 0.3M sodium acetate was added to each tube and eluted for at least 24 hours at 4°C. The eluate was transferred to a new 0.5 ml snap capped polypropylene microcentrifuge tube with a micropipette. A new RNAase free pipette tip was used for each tube and the RNA with ethanol was precipitated out.

***Ethanol Precipitation***

Two volumes of ice cold ethanol was added to each eluate then centrifuged at 15,000 xg for 15 minutes in a microcentrifuge. The supernatants were discharged and the precipitated RNA was re-dissolved in 100 µl of DEPC treated DI water. The RNA was stored in the microcentrifuge tubes at 4°C until needed. The tubes were stored at -20°C if storage was for more than two weeks. The following was order of rate of migration for each fragment in order from fastest to slowest:

- 5'CUAGAΨCUGGAGG3' (SEQ ID NO:1)
- Furan sided 5'CUPsoralenAGAΨCUGGAGG3' monoadducts (SEQ ID NO:2)
- 5'XXXXXXCCUCCAGAUAGXXXXXX3' (SEQ ID NO: 3)
- 5'XXXXXXCCUCCAGAUCCUPsoralenAGXXXXXX3' (SEQ ID NO: 4)

The tubes containing the remainder of each fraction were labeled and stored at -20°C.

#### **ETHANOL PRECIPITATION**

RNA oligonucleotide fragments were precipitated, and all glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." All solutions were stored in RNAase free glassware and introduction of nucleases was prevented. Absolute ethanol was stored at 0°C until used. Micropipettes were used to add two volumes of ice cold ethanol to nucleic acids that were to be precipitated in microcentrifuge tubes. Capped microcentrifuge tubes were placed into the microfuge and spun at 15,000 xg for 15 minutes. The supernatant was discarded and precipitated RNA was re-dissolved in DEPC treated DI-water. RNA was stored at 4°C in microcentrifuge tubes until ready to use.

#### **LIGATION OF RNA FRAGMENTS 2 AND 3**

All glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The following was added to a new 1.5 ml polypropylene snap capped microcentrifuge tube using a 100-1000  $\mu$ l pipette and a new sterile pipette tip was used for each solution:

Fragment 2 (3.0nM)	125.0 $\mu$ l
Fragment 3 (3.0nM)	125.0 $\mu$ l
Reaction buffer	250.0 $\mu$ l
RNA T4 ligase (9-12U/ml)	42 $\mu$ l

#### Reaction Buffer

RNase free DI-water	90.00ml
Tris-HCl (50mM)	0.79g
MgCl <sub>2</sub> (10mM)	0.20g
DTT (5mM)	0.078g
ATP (1mM)	0.55g
pH to 7.8 with HCL	
RNase free DI-water	QS to 100.00ml

The mixture was gently mixed and the RNA was melted by incubating the mixture at 16°C for one hour in a temperature controlled refrigerated chamber. RNA was precipitated out of the solution immediately after the incubation was completed.

#### *Alcohol Precipitation*

Two volumes (~1000  $\mu$ l) of ice cold absolute ethanol were added to the reaction mixture. The microcentrifuge tube was placed in a microcentrifuge at 15,000 xg for 15 minutes. The supernatant was decanted and discarded and the precipitated RNA was re-dissolved in 100  $\mu$ l DEPC treated water. The mixture was electrophoresed as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The following was the order of rate of migration for each fragment in order from fastest to slowest:

- a) Frag. 2  
5'CUAGAYCUGGAGG3'-OH<sub>Psoralen</sub> (SEQ ID NO: 5)

- b) Frag. 3  
5'UCCUGUGTΨCGAUCCACAGAAUUCGCACC-Puromycin  
(SEQ ID NO: 6)
- c) Frag 2+3  
Psoralen75'CUPsoralenAGAYCUGGAGGUCCUGUGTΨCGA  
UCCACAGAAUUCGCACC Puromycin (SEQ ID NO: 7)

Each fraction was isolated by UV shadowing, the bands were cut out, the RNAs were eluted from the gels and the RNA elute was precipitated out as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The ligation procedure was repeated with any residual unligated fragment 2 and 3 fractions. The ligated fractions 2 and 3 were pooled and stored in a small volume of RNase free DI-water at 4°C.

#### LIGATION OF RNA FRAGMENT 1 WITH FRAGMENT 2+3

All glassware was cleaned to remove any traces of RNase as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The following was added to a new 1.5 ml polypropylene snap capped microcentrifuge tube. A 100-1000 µl pipette and new tip was used for each solution:

Fragment 2+3 (3.0nM)	125.0µl
Reaction buffer	250.0µl
T4 Polynucleotide Kinase(5-10U/ml)	1.7µl
<u>Reaction Buffer</u>	
RNase free DI-water	90.00ml
Tris-HCl (40mM)	0.63g
MgCl <sub>2</sub> (10mM)	0.20g
DTT (5mM)	0.08g
ATP (1mM)	0.006g
pH to 7.8 with HCL	
RNase free DI-water	QS to 100.00ml

The RNA was gently mixed then melted by heating the mixture to 70°C for 5 minutes in a heating block. The mixture was cooled to room temperature over a two

hour period and the RNA was allowed to anneal in a tRNA configuration. The RNA was precipitated out of the solution.

#### *Alcohol Precipitation*

Two volumes (~1000 µl) of ice cold absolute ethanol were added to the reaction mixture. The microcentrifuge tube was placed in a microcentrifuge at 15,000 xg for 15 minutes. The supernatant was decanted and discarded and the precipitated RNA was re-dissolved in 100 µl DEPC treated water. The mixture was electrophoresed as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The following was the order of rate of migration for each fragment in order from fastest to slowest:

a) Frag. 1

5'GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACU3'  
(SEQ ID NO: 8)

b) Frag 2+3

Psoralen

5'CUPsoralenAGAYCUGGAGGUCCUGUGTΨCGAUCCACA  
GAAUUCGCACCPuromycin (SEQ ID NO: 6)

c) Frag. 1+2+3

Psoralen

5'GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUCUPsoralenAGAΨ  
CUGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCPuromycin (SEQ ID  
NO: 9)

Each fraction was isolated by UV shadowing, the bands were cut out, the RNAs were eluted from the gels and the RNA elute was precipitated out as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The ligation procedure was repeated with the unligated Fragment 1 and the 2+3 Fraction. The ligated fractions 2 +3 were pooled and stored in a small volume of RNase free DI-water at 4°C.

***Final RNA Ligation***

The following was added to a new 1.5ml polypropylene snap capped microcentrifuge tube. A 100-1000 µl pipette and new tip was used for each solution:

Fragment 1+2+3 (3.0nM)	250µl
reaction buffer	250µl
RNA T4 ligase (44µg/ml)	22µg

The mixture was incubated at 17°C in a temperature controlled refrigerator for 4.7 hours. Immediately after the incubation the tRNA was precipitated out as described in step 6.2 above and the tRNA was isolated by electrophoresis as described in the following section entitled "Polyacrylamide Gel Electrophoresis (PAGE) of Psoralenated RNA Fragments." The tRNA was pooled in a small volume of RNase free water and stored at 4°C for up to two weeks or stored at -20°C for periods longer than two weeks.

**POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE) OF PSORALENATED RNA FRAGMENTS*****Acrylamide Gel Preparation***

All reagents and glassware were made RNAase free as described in the following section entitled "Inactivation of RNases on Equipment, Supplies, and in Solutions." The gel apparatus was assembled to produce a 4 mm thick by 20 cm x 42 cm square gel. 29 parts acrylamide with 1 part ammonium crosslinker were mixed at room temperature with the appropriate amount of acrylamide solution in an RNAase free, thick walled Erlenmeyer flask.

Acrylamide Solution

urea (7M) 420.42 g

TBE (1X) QS to 1L

5X TBE

0.455 M Tris-HCl 53.9g

10mM EDTA 20ml of 0.5 M

RNAase free DI water 900ml

pH with boric acid to pH 9

QS with RNAase free DI water to 1L

The mixture was degassed with vacuum pressure for one minute. The appropriate amount of TEMED was added, mixed gently, and then the gel mixture was poured between the glass plates to within 0.5 cm of the top. The comb was immediately inserted between the glass sheets and into the gel mixture. An RNAase free gel comb was used. The comb produced wells for a 5 mm wide dye lane and 135 mm sample lanes. The gel was allowed to polymerize for about 30-40 minutes then the comb was carefully removed. The sample wells were rinsed out with a running buffer using a micropipette with a new pipette tip. The wells were then filled with running buffer.

*Sample Preparation*

An aliquot of the sample was suspended in loading buffer in a snap capped microcentrifuge tube and vortex mixed. Indicator dye was not added to the sample.

Loading Buffer

Urea (7M) 420.42 g

Tris HCl (50mM) 7.85 g

QS with RNAase free D-H<sub>2</sub>O to 1L*Electrophoresis run*

The maximum volume of RNA/loading buffer solution was loaded into the 135 mm sample wells and the appropriate volume of tracking dye in 5 mm tracking lane. The samples were electrophoresed in a 5°C refrigerator. The electrophoresis was stopped when the tracking dye reached the edge of the gel. The apparatus was then disassembled. Glass panels were not removed from the gel. The gel-glass panel unit

was placed on a UV light box. With UV filtering goggles in place, the UV lights were turned on. The RNA bands were identified. They appeared as denser shadows under UV lighting conditions. The RNA was extracted from the gel. Each band was excised with a new sterile and RNAase free scalpel blade and each band was transferred into a new 1.5 ml snap capped microcentrifuge tube. Each gel was crushed against the walls of the microcentrifuge tubes with the side of the scalpel blade. A new blade was used for each sample. 1.0 ml of 0.3M sodium acetate was added to each tube and eluted for at least 24 hours at 4°C. The eluate was transferred to a new 0.5 ml snap capped polypropylene microcentrifuge tubes with a micropipette with a new RNAase free pipette tip for each tube. Two volumes of ice cold ethanol was added to each eluate, then centrifuged at 15,000 xg for 15 minutes in a microcentrifuge. The supernatants were discarded and the precipitated RNA was redissolved in 100 µl of DEPC treated DI water. The RNA was stored in the microcentrifuge tubes at 4°C until needed.

#### **OLIGONUCLEOTIDE ANALYSIS BY HPLC**

HPLC purification of the RNA oligonucleotides was performed using anion exchange chromatography. Either the 2'-protected or 2'-deprotected forms may be chromatographed. The 2'-protected form offered the advantage of minimizing secondary structure effects and providing resistance to nucleases. If the RNA was fully deprotected, sterile conditions were required during purification.

One skilled in the art will understand that the HPLC purification methods of Example 2 may be modified in order to purify the RNA oligonucleotides. Modification of the HPLC purification methods of Example 2, including HPLC gradient, temperature, and other parameters, may be necessary. One of skill in the art would also recognize that a one-step HPLC purification method may also be used in accordance with several embodiments of the current invention.

#### **INACTIVATION OF RNASES ON EQUIPMENT, SUPPLIES, AND IN SOLUTIONS**

Glassware was treated by baking at 180°C for at least 8 hours. Plasticware was treated by rinsing with chloroform. Alternatively, all items were soaked in 0.1% DEPC.

##### ***Treatment with 0.1% DEPC***



0.1% DEPC was prepared. DI water was filtered through a 0.2 $\mu$ M membrane filter. The water was autoclaved at 15 psi for 15 minutes on a liquid cycle. 1.0g (wt/v) DEPC/liter of sterile filtered water was added.

#### *Glass and Plasticware*

All glass and plasticware was submerged in 0.1% DEPC for two hours at 37°C. The glassware was rinsed at least 5X with sterile DI water. The glassware was heated to 100°C for 15 minutes or autoclaved for 15 minutes at 15 psi on a liquid cycle.

#### *Electrophoresis Tanks Used for Electrophoresis of RNA*

Tanks were washed with detergent, rinsed with water then ethanol and air dried. The tank was filled with 3% (v/v) hydrogen peroxide (30ml/L) and left standing for 10 minutes at room temperature. The tank was rinsed at least 5 times with DEPC treated water.

#### *Solutions*

All solutions were made using RNase free glassware, plastic ware, autoclaved water, chemicals reserved for work with RNA and RNase free spatulas. Disposable gloves were used. When possible, the solutions were treated with 0.1% DEPC for at least 12 hours at 37°C and then heated to 100°C for 15 minutes or autoclaved for 15 minutes at 15 psi on a liquid cycle.

#### **RNA TRANSLATION**

2  $\mu$ l of gastroinhibitory peptide (GIP) mRNA at a concentration of 20  $\mu$ l/ml was placed in a 250  $\mu$ l snapcap polypropylene microcentrifuge tube. 35  $\mu$ l of rabbit reticulocyte lysate (available commercially from Promega) was added. 1  $\mu$ l of amino acid mixture which did not contain methionine (available commercially from Promega) was added. 1  $\mu$ l of  $^{35}$ S methionine or unlabeled methionine was added. 2  $\mu$ l of  $^{32}$ P GIP mRNA or unlabeled GIP mRNA was added. Optionally, 2 ml of luciferase may be added to some tubes to serve as a control. In a preferred embodiment, luciferase was used instead of GIP mRNA. One skilled in the art will understand that indeed any mRNA fragment containing the appropriate sequences may be used.

SATA was added to the experimental tubes. Control tubes which did not contain SATA were also prepared. The quantity of SATA used was approximately between 0.1 $\mu$ g to 500  $\mu$ g, preferably between 0.5  $\mu$ g to 50  $\mu$ g. 1  $\mu$ l of Rnasin at 40 units/ml was added. Nuclease free water was added to make a total volume of 50  $\mu$ l.

For proteins greater than approximately 150 amino acids, the amount of tRNA may need to be supplemented. For example, approximately 10 - 200  $\mu$ g of tRNA may be added. In general, the quantity of the SATA should be high enough to effectively suppress stop or pseudo stop codons. The quantity of the native tRNA must be high enough to out compete the SATA which does not undergo dynamic proofreading under the action of elongation factors.

Each tube was immediately capped, parafilmed and incubated for the translation reactions at 30°C for 90 minutes. The contents of each reaction tube was transferred into a 50  $\mu$ l quartz capillary tube by capillary action. The SATA was crosslinked with mRNA by illuminating the contents of each tube with 2-10 J/cm<sup>2</sup> ~350 nm wavelength light, as per Gasparro *et al.* (Photochem. Photobiol. 57:1007 (1993), herein incorporated by reference). Following photocrosslinking, the contents of each tube were transferred into a new snapcap microfuge tube. The ribosomes were dissociated by chelating the calcium cations by adding 2  $\mu$ l of 10 mM EDTA to each tube. Between each step, each tube was gently mixed by stirring each component with a pipette tip upon addition.

The optimal RNA for a translation was determined prior to performing definitive experiments. Serial dilutions may be required to find the optimal concentration of mRNA between 5-20  $\mu$ g/ml.

Reagent	1	2	3	4
Rabbit reticulocyte lysate (35 $\mu$ l)	+	+	+	+
Amino acid mixture minus methionine (1 $\mu$ l of 1 mM)	+	+	+	+
<sup>35</sup> S Methionine (1 $\mu$ l of 1,200Ci/mmol)	+	-	-	+
Methionine (unlabeled)	-	+	+	-
GIP mRNA (2 $\mu$ l of 20 $\mu$ g/ml)	+	-	-	-
<sup>32</sup> P GIP mRNA (2 $\mu$ l of 20 $\mu$ g/ml)	-	+	+	-
Rnasin (1 $\mu$ l of 40 U/ $\mu$ l)	+	+	+	+
SATA	-	-	-	-
Water, nuclease free (q.s. to 50 $\mu$ l)	+	+	+	+

SDS-Page electrophoresis was performed on each sample, as described above. Autoradiography on the gel was performed, as described by Sambrook et. al., Molecular Cloning, A Laboratory Manual, 2<sup>nd</sup> ed., Coldspring Harbor Press (1989), herein incorporated by reference.

The above example teaches the production and use of SATA (e.g., puromycin on tRNA plus crosslinker on the tRNA) and the production and use of Linking tRNA Analog (e.g., no puromycin, but has crosslinker on tRNA).

In another example, the SATA was produced in a manner similar to the above methodology, except that uridines were substituted with pseudouridines. Substitution by pseudouridines can also be used with Linking tRNA Analog, as it facilitates the formation of crosslinker monoadduct formation (such as formation of the psoralen monoadduct). This technique is discussed below in Example 2.

#### **EXAMPLE 2: PRODUCTION OF THE SATA USING PSEUDOURIDINE**

As discussed above, one skilled in the art will appreciate that the SATA, Linking tRNA Analog and Nonsense Suppressor tRNA can be produced in a number of different ways. Figure 5 shows the chemical structures for uridine and pseudouridine. Pseudouridine is a naturally occurring base found in tRNA that forms hydrogen bonds just as uridine does, but lacks the 5-6 double bond that is the target for psoralen. Pseudouridine, as used herein, shall include the naturally occurring base and any

synthetic analogs or modifications. In a preferred embodiment, the SATA was produced using pseudouridine. Linking tRNA Analog can also be produced using pseudouridine. Specifically, in a preferred embodiment, three fragments (Fig. 1) were purchased from a commercial source (Dharmacon Research Inc., Boulder, CO). Modified bases and a fragment 3 ("Fragment 3") with a pre-attached puromycin on its 3' end and a PO<sub>4</sub> on its 3' end were included, all of which are available commercially. The three fragments were used to facilitate manipulation of a fragment 2 ("Fragment 2") in forming the monoadduct. Sequences of the three fragments, according to some embodiments, are as follows (2 example sequences are provided for each fragment):

*Fragment 1*

5'PO<sub>4</sub>GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACOH3' (SEQ ID NO: 10)

5'PO<sub>4</sub>GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACOH3' (SEQ ID NO: 16)

*Fragment 2*

5'OHΨCUAACΨCOH3' (SEQ ID NO: 11)

5' OHΨCUAAAΨCOH 3' (SEQ ID NO: 17)

*Fragment 3*

5'PO<sub>4</sub>UGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCPuromycin3' (SEQ ID NO: 12)

5'PO<sub>4</sub>UGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCPuromycin3' (SEQ ID NO: 18)

The above sequences listed in Fragment 3 are applicable for SATA. For Linking tRNA Analogs, the sequences would be similar, except the puromycin would be replaced by adenosine.

Modified yeast tRNA<sup>Ala</sup> or yeast tRNA<sup>Phe</sup> was used according to one embodiment of the invention. However, one skilled in the art will understand that sequences can be chosen widely from known tRNAs or by selecting sequences that will form into a tRNA-like structure. One advantage of using pseudouridine in some embodiments is that the pseudouridine in Fragment 2 avoids psoralen labeling of the nontarget U's. Use of pseudouridine instead of uridine decreases the avidity of the A

site of the ribosome for the tRNA analog but eliminates the interaction of the terminal uridine with psoralen. The use of the Yarus "extended anticodon" guidelines increases A site binding (Yarus, Science 218:646-652, 1982, herein incorporated by reference).

In one embodiment, Fragment 2 was used in a helical conformation to induce the psoralen to intercalate. One skilled in the art will understand that other conformations can also be used in accordance with several embodiments of the invention. A complementary strand was also used. RNA or DNA was used, and a sequence, such as poly C or poly G when C interacts with the psoralen to one or both ends, was added to facilitate separation and removal after monoadduct formation was accomplished. Use of pseudouridine instead of uridines in the complement permitted the use of a high efficiency wave length, such as about 365 nm, without fear of crosslinking the product. Irradiation was preferably in the range of about 300-450 nm, more preferably in the range of about 320 to 400 nm, and most preferably about 365 nm. Further, use of pseudouridine left the furan-sided monoadduct in place on Fragment 2 because the Maf is the predominate first step in the crosslink formation.

The following crRNA sequences with pseudouridine were used according to a preferred embodiment of the present invention. One skilled in the art will understand that substitutions and modifications of these sequences, and of the other sequences listed herein, can also be used in accordance with several embodiments of the current invention. For example, for SEQ ID NO: 19, listed below, the sequence can also be 5'XXXXXXGAYΨΨAGAXXXXXXX3' (SEQ ID NO: 30):

CCCΨCCAGAGΨΨAGACCC (SEQ ID NO: 13)

5'CCCCCGAYΨΨAGACCCCCC3' (SEQ ID NO: 19)

#### **Step 1: Furan Sided Monoadduction Of Psoralen To Fragment 2**

The formation of a furan sided psoralen monoadduct with the target uridine of Fragment 2 was performed as follows:

A reaction buffer was prepared as follows:

Tris HCL	25 mM
NaCl	100 mM
EDTA	0.32 mM
pH	7.0

4'-hydroxy methyl-4,5',8'-triethyl psoralen (HMT) was then added to a final concentration of 0.32 mM and equimolar amounts of fragment 2 and cRNA were added to a final molar ratio of fragment 2: cRNA : psoralen = 1:1:1000. A total volume of 100 $\mu$ l was irradiated at a time.

The mixture of complementary oligos, HMT, psoralen was processed as follows:

1) Heated to 85° C for 60 sec followed by cooling to 4° C over 15 min, using PCR thermocycler.

2) Irradiated for 20 min at 4° C, in Eppendorf UVette plastic cuvette, covered top with parafilm, laid on the top of UV lamp (1mW/cm<sup>2</sup> multi-wavelength UV lamp ( $\lambda$ >300nm) (UV L21 model  $\lambda$  365 nm).

Steps 1 and 2 above were repeated 4 times to re-intercalate and irradiate HMT. After the second irradiation additional 10  $\mu$ l of 1.6 mM HMT was added in total 100 $\mu$ l reaction volume. After 4 cycles of irradiation, the free psoralens were extracted with chloroform and all oligos (labeled and unlabeled) were precipitated with ethanol overnight (see precipitation step). A small aliquot was saved for gel identification.

#### **Step 2: Purification Of HMT Conjugated Fragment 2 (2MA) Oligo By HPLC**

1) The reaction mixture was dried with speed vacuum for 10 minutes and then was dissolved with 2  $\mu$ l of 0.1 M TEAA, pH 7.0 buffer.

0.1 M TEAA, pH 7.0 Buffer

Acetic Acid            5.6 ml

Triethylamine        13.86 ml

H<sub>2</sub>O (RNAase free)   950 ml

pH adjusted to 7.0 with acetic acid

and water added to 1L

2) The sample was loaded onto a Waters Xterra MS C18, 2.5  $\mu$ m, 4.5x50 mm reverse-phase column pre-equilibrated with buffer A (5% wt/wt acetonitrile in 0.1M TEAA, pH 7.0) The sample was eluted with a gradient of 0-55% buffer B (15% wt/wt acetonitrile in 0.1M TEAA, pH 7.0) to buffer A over a 35 minute time frame at a flow rate of 1 ml/minute. The column temperature was 60°C and the detection wave length, set by a narrow band filter, was 340 nm. Furan sided psoralen

monoadduct absorbs at 340 nm but the RNA, and any pyrone sided monoadduct does not. The buffer solutions were filtered and degassed before use.

The 2MA eluted at around 25-28 minutes at a buffer B concentration of 40%. Unpsoralenated fragment 2 eluted before 8 minutes based on subsequent gel electrophoresis analysis on collected fractions.

The column was washed with 100% acetonitrile for 5 minutes and was re-equilibrated with buffer A for 15 minutes. All fractions were dried with speed vacuum overnight.

The fractions containing the 2MA were identified by the level of absorbance at 260 nm (RNA) and 330 nm (furan sided psoralen monoadducted RNA). This was done by redissolving the dried fractions with 120 µl of Rnase-free distilled water and the absorbance was measured with a spectrophotometer at 260 nm and 330 nm. The fractions with high absorbance at both wavelengths were pooled then dried with speed vacuum. A small aliquot from each was saved for gel analysis.

The cross-linked products were analyzed on a denaturing 20% TBE-urea gel and visualized by gel silver staining.

### **Step 3: Purification Of HMT Conjugated Fragment 2 Oligo From cRNA By HPLC**

The dried samples were pooled and then were dissolved with 0.5X TE buffer. A sample of about 0.4 absorbance unit was loaded onto a Dionex DNAPac PA-100 (4x250mm) column which was pre-equilibrated with buffer C (25 mM Tris-HCl, pH 8.0) and the column temperature was 85°C (anion exchange HPLC).

The oligos were eluted at a flow rate of 1 ml/min. with a concave gradient from 4% to 55% buffer D for 15 minutes followed by a convex gradient from 55 % to 80% with buffer D for the next 15 minutes. The oligos were washed with 100% buffer D for 5 min and 100% buffer C for another 5 min at a flow rate of 1.5 ml/min; Fractions were collected that absorbed 260 nm light. 2MA had a retention time (RT) of 16.2 minutes and was eluted by 57% buffer D, and free fragment 2 had RT less than 16.6 minutes, and was eluted by 55% buffer D and free cRNA had RT greater than 19.2 minutes. The fractions were collected that absorbed at 254 or 260 nm. The collected fractions were dried with speed vacuum overnight. All solutions were filtered and degassed before use.

The solution used comprised the following:

C: 25mM Tris-HCl pH 8.0;

D: 250mM NaClO<sub>4</sub> in 25mM Tris pH 8.0 buffer.

TE: 10mM Tris-HCl pH 8.0 with 1mM EDTA

**Step 4: Desalting, Precipitation And Collection Of The Purified 2MA Oligo**

The dried fractions were redissolved with 100µl Rnase free distilled water. 500µl cool 100% ethanol with 0.5M (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> was added and the mixture was vortexed briefly. The mixture was then frozen on dry ice for 60 minutes or stored at -20°C overnight.

The samples were then brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant was decanted or removed by pipette. Care was taken not to disturb pellet. If the pellet still contained salt, this step was repeated. The pellet was then washed with 70% pre-cooled ethanol twice. The wet pellet was dried with speed vacuum for 15 min. Urea PAGE gel identified the right fractions for the next step.

**Step 5: Ligation Of 2MA Oligo To Fragment 3 Oligo**

The following steps were performed:



***A. The following reagents and instruments were used:***

Nuclease-Free Water (Promega )  
 polyethylene glycol (PEG8000 Sigma) 40%(wt/wt in water)  
 RNasin ® Ribonuclease Inhibitor (Promega)  
 phenol:chloroform  
 1.5 ml sterile microcentrifuge tubes  
 100% ethanol  
 70% ethanol  
 Dry ice or -20°C freezer  
 Microcentrifuge at room temperature and +4°C  
 PCR thermocycler or water bath

***B. The following reaction conditions were used:***

50 mM Tris-HCl (pH 7.8)  
 10 mM MgCl<sub>2</sub>,  
 10 mM DTT  
 1 mM ATP  
 18-20% PEG

***C. The following reaction mixture was assembled in a sterile microcentrifuge tube:***

Fragment 3 (Donor) 1µl (6µg) (Purified, when necessary, before using as a donor)  
 2MA (Acceptor) 1µl (1.5µg).

After adding 8 µl Rnase free dH<sub>2</sub>O 8µl, the reactions were incubated at 85° C for 1 minute to relax the oligo secondary structure, then slowly cooled to 4° C, using a PCR machine thermocycler. The preheated tube was placed on ice to keep cool and centrifuged briefly, then the following was added:

10X Ligase Buffer	4µl
10mM ATP	4µl
Rnase Out or Rnasin(40u/µl) Promega	0.5µl
PEG, 40 % (Sigma)	20µl
T4 RNA Ligase (10u/µl) (NEB)	1µl

Nuclease-free water was added to final Volume of 40  $\mu$ l. The mixture was incubate at 16° C overnight(16hr). The mixture was centrifuged briefly and then was placed on ice.

***D. Precipitation of Oligonucleotides:***

60 $\mu$ l DEPC RNase.free distilled water was added to the mixture and then 150  $\mu$ l phenol/chloroform was added. The mixture was vortexed vigorously for 30 seconds. The precipitate was then centrifuged out at maximum speed in a microcentrifuge for 5 minutes at room temperature. The aqueous phase was transferred to a new microcentrifuge tube (>95 $\mu$ l).

To this was added 3 $\mu$ l 5 mg/ml glycogen, and 500 $\mu$ l pre-cooled 100% ethanol with 0.5M (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> and the mixture was vortexed briefly and then was frozen on dry ice for 60 minutes. At this point, it may be stored overnight at -20°C. The dried fractions were redissolved with 100 $\mu$ l RNase-free distilled water, 500 $\mu$ l cool 100% ethanol with 0.5M (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> was added and vortexed briefly. This was then frozen on dry ice for 60 minutes or stored at -20C overnight. The samples were then brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes and supernatant removed by pipette. Care was taken not to disturb pellet. If the pellet still contained salt, this step was repeated once. The pellet was then washed with 70% pre-cooled ethanol several times. This was then centrifuged at maximum speed in a microcentrifuge for 5 minutes at 4C. The ethanol was carefully removed using a pipette. Centrifugation was repeated again to collect remaining ethanol which was carefully removed. The wet pellet was dried with speed vacuum for 10 min. A small aliquot was collected for the gel analysis. For long term storage, the RNA was stored in ethanol at -20C. Care was taken not to store the RNA in DEPC water.

**Step 6: Purification Of The Ligated Fragment 3 Oligo Complex**

The dried sample was redissolved with 0.5X TE buffer and was loaded onto a DNAPac PA-100 column which was equilibrated with buffer C. The column temperature was 85°C and the detector operated at 254 nm to identify fractions with RNA and at 340 nm to identify fractions with 2MaF. The oligos were eluted with a convex gradient from 30%to 70% with buffer D for the first 20 minutes at a flow rate of 0.8ml/min and followed with a linear gradient from 70 % to 98% D for another 20 min at the same flow rate. The elution was completed by washing with 100% D for 7

min and 100% C for another 10 min at 1.0 ml/min flow rate. The fractions were detected with 254 or 260 nm wavelength light. The ligated oligos (2MA-fragment 3) were eluted after 34 min, by more than 90% buffer B. Fractions with 254 nm absorbance ( $A_{254nm} > 0.01$ ) were collected and dried with speed vacuum overnight.

#### **Step 7: Purified 2MA-Fragment 3 Desalting And Precipitation**

The dried fractions were re-dissolved with 100 $\mu$ l Rnase free distilled water, 500 $\mu$ l cool 100% ethanol with 0.5M (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> was added and the mixture was vortexed briefly. The mixture was then frozen on dry ice for 60 minutes or stored at -20C overnight.

The samples were brought to 4° C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant decanted or removed by pipette. Care was taken not to disturb pellet. If still containing salt, this step was repeated. The pellet was then washed with 70% pre-cooled ethanol twice. The wet pellet was dried with speed vacuum for 15 min.

Urea PAGE was performed to identify the ligated 2MA-fragment-3 for use in the next step of ligating fragment 1 to the 2MA-fragment-3 oligo which completes the SATA linker.

#### **Step 8: Preparation Of SATA (Or Other tRNA Molecule)**

##### ***A. RNA Oligo 5'phosphorylation***

##### ***1. Reagent and instrument:***

- Nuclease-Free Water (Cat.# P1193 Promega)
- RNasin ® Ribonuclease Inhibitor (Cat# N2511 Promega)
- Phenol:chloroform
- Sterile microcentrifuge tubes
- 100% ethanol
- 70% ethanol
- Microcentrifuge at room temperature and 4°C
- PCR thermalcycler or water bath

##### ***2. Assemble the following reaction mixture in a sterile microcentrifuge tube:***

<u>Component</u>	<u>Volume</u>
• Acceptor RNA	<200ng
• T4 ligase 10X Reaction Buffer*	4 $\mu$ l

- RNasin ® Ribonuclease Inhibitor (40u/μl)      20unit
- T4 kinase (9-12u/μl)      2μl
- 10 mM ATP      4μl
- Nuclease-Free Water to final volume      40μl

Incubate at 37°C for 30 minutes in a PCR thermocycler or water bath. For non-radioactive phosphorylation, use up to 300 pmol of 5' termini in a 30 to 40 μl reaction containing 1X T4 Polynucleotide Kinase Reaction Buffer, 1 mM ATP and 10 to 20 units of T4 Polynucleotide Kinase. Incubate at 37°C for 30 minutes. 1X T4 DNA Ligase Reaction Buffer contains 1 mM ATP and can be substituted in non-radioactive phosphorylations. T4 Polynucleotide Kinase exhibits 100% activity in this buffer). Fresh buffer is required for optimal activity (in older buffers, loss of DTT due to oxidation lowers activity).

***B. Annealing Fragment1 and 2MA-fragment 3 oligo complex:***

***1. Reagents and instruments:***

- PCR thermocycler instrument or water bath
- 100μg/ml nuclease-free albumin
- 100mM MgCl<sub>2</sub>

***2. Assemble the following reaction mixture in a sterile microcentrifuge tube:***

- Acceptor RNA oligo (1E)      <200 ng
- Donor RNA oligo (3G-2G ligated oligo)      <200 ng
- (5' phosphorylated oligo from step A)

Appropriate ratios are as follows: Acceptor oligo:Donor oligo (Fragment 1: 2MA-Fragment 3) molar ratio should be 1:1.1 to avoid fragment 1 self-ligation. MgCl<sub>2</sub> was added to T4 ligase buffer (50mM Tris-HCl<sub>4</sub>-(pH 7.8), 10 mM MgCl<sub>2</sub>, 10 mM DTT and 1mM ATP) to final 20 mM concentration. Add Rnase free albumin to final 5μg/ml. The final volume should be no more than 100μl. The solution was heated to 70° C for 5 min, then was cooled from 70° C to 26° C over 2 hours and cooled from 26°C to 0°C over 40 minutes. Incubate at 16°C for 16 to 17 hours using PCR instrument.

***C. Ligation of annealed oligos***

- Annealed oligos      < 15μl
- 10mM ATP      2μl

- |  |       |
|--|-------|
| • 40% PEG                                  | 18μl  |
| • T4 ligase 10X Buffer                     | 2μl   |
| • RNasin ® Ribonuclease Inhibitor (40u/μl) | 0.5μl |
| • T4 ligase (9-12u/μl)(NEB)                | 1μl   |
| • Nuclease-Free Water to final volume      | 40μl  |

#### ***D. Precipitating tRNA fragment***

After ligation, 50 μl DEPC water and 150 μl phenol: chloroform were added and vortexed vigorously for 30 seconds. This was then centrifuged at maximum speed in a microcentrifuge for 5 minutes at room temperature. The aqueous phase was transferred to a new microcentrifuge tube (~100 μl). To this was added 2 μl 10mg/ml mussel glycogen, 10μl 3M sodium acetate, pH 5.2. This was mixed well. Then 220μl 95% ethanol was added and vortexed briefly. The mixture was then frozen on dry ice for 30 minutes. At this point the mixture may be stored over night at -20°C or one may proceed. In one embodiment, the RNA should preferably not be stored in DEPC water, but in ethanol, at -20° C.

Then the samples were brought to 4°C and centrifuged at maximum speed in a microcentrifuge for 15 minutes. The position of the pellet was noted and the supernatant decanted or removed by pipette. Care was taken not to disturb pellet. The pellet was then washed with 70% pre-cooled ethanol twice. After removing the ethanol, the wet pellet was dried with a speed vacuum for 15 min. The dried pellet was stored at -20°C, until the next step.

#### **RNA Translation**

A luciferase mRNA which was modified to have the stop codon corresponding to that recognized by the anticodon of the SATA ( in the present case UAG) was used in a standard Promega *in vitro* translation kit in the recommended 1μl of concentration 1μg/μl. One skilled in the art will understand that indeed any mRNA fragment containing the appropriate sequences may be used.

SATA was added to the experimental tubes. Control tubes which did not contain SATA were also prepared. The quantity of SATA used was approximately between 0.1μg to 500 μg, preferably between 0.5 μg to 50 μg. 1 μl of Rnasin at 40 units/ml was added. Nuclease free water was added to make a total volume of 50 μl.

For proteins greater than approximately 150 amino acids, the amount of tRNA may need to be supplemented. For example, approximately 10 - 200  $\mu\text{g}$  of tRNA may be added. In general, the quantity of the SATA should be high enough to effectively suppress stop or pseudo stop codons. The quantity of the native tRNA must be high enough to out compete the SATA which does not undergo dynamic proofreading under the action of elongation factors.

Each tube was immediately capped, parafilmed and incubated for the translation reactions at 30°C for 90 minutes. The contents of each reaction tube was transferred into a 50  $\mu\text{l}$  quartz capillary tube by capillary action. The SATA was crosslinked with mRNA by illuminating the contents of each tube with 2-10 J/cm<sup>2</sup> ~350 nm wavelength light, as per Gasparro *et al.* (Photochem. Photobiol. 57:1007 (1993), herein incorporated by reference). Following photocrosslinking, the contents of each tube were transferred into a new snapcap microfuge tube. The ribosomes were dissociated by chelating the calcium cations by adding 2  $\mu\text{l}$  of 10 mM EDTA to each tube. Between each step, each tube was gently mixed by stirring each component with a pipette tip upon addition.

The optimal RNA for a translation was determined prior to performing definitive experiments. Serial dilutions may be required to find the optimal concentration of mRNA between 5-20  $\mu\text{g}/\text{ml}$ .

SDS-Page electrophoresis was performed on each sample, as described above. Autoradiography on the gel was performed, as described by Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2<sup>nd</sup> ed., Cold Spring Harbor Press (1989), herein incorporated by reference.

The above example is instructive for the production and use of SATA (puromycin on tRNA and crosslinker on tRNA) and for the production and use of Linking tRNA Analog (no puromycin, with crosslinker on tRNA).

**EXAMPLE 3: PRODUCTION OF LINKING tRNA ANALOG USING  
RIBONUCLEOTIDES MODIFIED TO FORM CROSSLINKERS: USE OF  
PSORALEN AND NON-PSORALEN CROSSLINKERS**

As described above, pseudouridine can be used in some embodiments to minimize the formation of unwanted monoadducts and crosslinks. In one embodiment, a crosslinker modified mononucleotide is formed and used. One advantage of the

crosslinker modified mononucleotide is that it minimizes the formation of undesirable monoadducts and crosslinks.

As discussed above, one skilled in the art will appreciate that the SATA, Linking tRNA Analog, and Nonsense Suppressor Analog can be produced in a number of different ways. In a preferred embodiment, psoralenated uridine 5' mononucleotide, 2-thiocytosine, 2-thiouridine, 4-thiouridine 5-iodocytosine, 5-iodouridine, 5-bromouridine or 2-chloroadenosine can be produced or purchased and enzymatically ligated to an oligonucleotide to be incorporated into a tRNA analog. Aryl azides, and analogues of aryl azides, and any modifications thereto, can also be used in several embodiments, as a linking moiety or agent. The following protocol can be employed for crosslinkers that are located on the tRNA. One skilled in the art will understand that this protocol can also be used for crosslinkers located on the mRNA. Thus, the following example is instructive on the production and use of SATA, Linking tRNA Analog, and Nonsense Suppressor Analog.

#### **Production Of Modified Nucleotide**

4-thioU, 5-iodo and 5-bromo U with and without puromycin can be purchased already incorporated into a custom nucleotide up to 80 basepairs in length (Dharmacon, Inc). Therefore, the SATA, and the Linking tRNA Analog with these crosslinkers already in place, and similar crosslinkers, can be purchased directly from Dharmacon, Inc. Nonsense Suppressor Analog can also be purchased from Dharmacon, Inc.

2-thiocytosine, 2-thiouridine, 4thiouridine 5-iodocytosine, 5-iodouridine, 5-bromouridine or 2-chloroadenosine can all be purchased for crosslinking from Ambion, Inc. for the use in the Ambion MODIscrip kit for incorporation into RNA. Therefore, the SATA and the Linking tRNA Analog along with these crosslinkers, and similar crosslinkers, can be purchased directly from Ambion, Inc

The  $\text{PO}_4\text{U}_{\text{psoralen}}$  can be produced as follows:

AUAUAUAUAUAUAUAUAUAUGGGGGG (seq A1) (SEQ ID NO: 20)  
(available from Dharmacon, Inc.)

CCCCCATATATATATATATATATAT (seq A2) (SEQ ID NO: 21)  
(available from University of Southern California services).

The formation of a furan-sided psoralen monoadduct with the target uridine is performed as follows:

A reaction buffer is prepared. The reaction buffer, with a pH of 7.0, contains 25 mM Tris HCL, 100 mM NaCl, and 0.32 mM EDTA.

4'-hydroxy methyl-4,5',8'-triethyl psoralen (HMT) is then added to a final concentration of 0.32 mM and equimolar amounts of seq A1 and seq A2 are added to a final molar ratio of seq A1: seq A2 : psoralen = 1:1:1000. A total volume of 100µl is irradiated at a time.

The mixture of complementary oligos, HMT, trimethylpsoralen is processed as follows: 1) Heat to 85° C for 60 sec followed by cooling to 4° C over 15 min, using PCR thermocycler; and 2) Irradiate for 20 to 60 min at 4° C, in Eppendorf UVette plastic cuvette, covered top with parafilm, in an RPR-200 Rayonet Chamber Reactor equipped with a cooling fan and 419 nm wave. This is either placed on an ice water bath or in a -20° C freezer.

Steps 1 and 2 above are repeated 4 times to re-intercalate and irradiate HMT. After 4 cycles of irradiation, the free psoralens are extracted with chloroform and all oligos (labeled and unlabeled) are precipitated with ethanol overnight (see precipitation step). A small aliquot is saved for gel identification.

Comparable sequences can be produced using the Ambion, Inc kit for non-psoralen crosslinkers.

#### **RNase H Digestion Of RNAs In DNA/RNA duplexes**

The following steps are performed: (1) Dry down oligos in speed vac; (2) Resuspend pellet in 10 µL 1X Hyb Mix; (3) Heat at 68°C for 10 minutes; (4) Cool slowly to 30°C. Pulse spin down; (5) Add 10 µL 2X RNase H Buffer. Mix. (6) Incubate at 30°C for 60 minutes; (7) Add 130 µL Stop Mix.

For the Phenol/Chloroform extract: (1) Add 1 vol. phenol/chloroform; (2) Vortex well; (3) Spin down 2 minutes in room temperature microfuge; (4) Remove top layer to new tube.

For the Chloroform extract: (1) Add 1 vol. chloroform; (2) Vortex well; (3) Spin down 2 minutes in room temperature microfuge; (4) Remove top layer to new tube.

Then, (1) Add 375 µL 100% ethanol; (2) Freeze at -80°C; (3) Spin down 10 minutes in room temperature microfuge; (4) Wash pellet with 70% ethanol; (5) Resuspend in 10 µL loading dye; (6) Heat at 100°C for 3 minutes immediately before loading.



Purification of monoribonucleotides nucleotides from the longer cDNA as well as longer RNA fragments, is accomplished using anion exchange HPLC. The psoralen-monoadducted mononucleotides ( $\text{PO}_4\text{U}_{\text{psoralen}}$ ) are then separated by reverse phase HPLC from mononucleotides that were not monoadducted ( $\text{PO}_4\text{U}$  and  $\text{PO}_4\text{A}$ ).

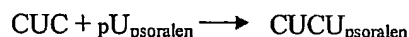
Similar digestion techniques and nucleotide incorporation, described below, can also be used for non-psoralen crosslinkers using the Ambion, Inc kit.

### **Incorporation Of Light Sensitive Nucleotides Into The TRNA Component Oligoribonucleotides**

The following protocol can be used for incorporating a  $\text{pU}_{\text{crosslinker}}$  into a CUA stop anticodon. However, one skilled in the art will understand that other nucleotides can also be used to produce other stop anticodons and pseudo stop anticodons in accordance with the methods described herein

Generally, methods adapted from the protocols for T4 RNA ligase are used, but with some modification because of the lack of protection of the 3' OH of the modified nucleotides.

5'OH CUC OH 3' oligoribonucleotides (seq B1) can be purchased from Dharmacon, Inc. and can be as acceptors in the ligation. The molar ratio of B1 to psoralenated mononucleotides is preferably kept at 10:1 to 50:1 so that the modified U's will be greatly out-numbered, thereby preventing the formation of  $\text{CUC}(\text{U}_{\text{crosslinker}})_N$ . This makes one of the preferred reactions:



In one embodiment, the product is purified by sequential anion exchange and reversed phase HPLC to ensure that the psoralenated U and the longer psoralenated 7 mer are separated. The 7 mer is then 3' protected by ligation with pAp yielding  $\text{CUCU}_{\text{crosslinker}}\text{Ap}$  (Fragment 2B).

This is again purified with anion exchange HPLCF or the next ligation.

### **First Ligation Of Fragment 2B To 1B Or 1B1**

This 2B fragment can be used in a tRNA analog that has a stable acceptor or one that has a native esterified acceptor. In one embodiment, to assure that the native 3' end can be aminoacylated by native AA-tRNA synthetases, the acceptor stem is modified in that version of the analog. In the SATA version, in one embodiment, the 3'

fragment is maintained with a commercially prepared puromycin as the acceptor. Thus, in one embodiment, the following are used in two different 5' ends:

5' OHGCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA 3' seq 1B (SEQ ID NO: 22) (to be used with the tRNA analog with the stable puromycin acceptor) and

5' OHGGGGCUUUAGCUCAGUUGGGAGAGCGCCAGA 3' seq 1B<sub>1</sub> (SEQ ID NO: 23) (to be used with the native esterified acceptor).

The ligation is performed again with T4 RNA ligase and purified by length.

The equation for sequence 1B is as follows:

5' OHGCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA 3' + CUCU<sub>crosslinker</sub>APO<sub>4</sub> 3' (SEQ ID NO: 22)  $\longrightarrow$

5' OHGCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUCU<sub>crosslinker</sub>APO<sub>4</sub> 3' (SEQ ID NO: 24)

For sequence 1B<sub>1</sub>:

5' OHGGGGCUUUAGCUCAGUUGGGAGAGCGCCAGA + CUCU<sub>crosslinker</sub>APO<sub>4</sub> 3' (SEQ ID NO: 23)  $\longrightarrow$

5' OHGGGGCUUUAGCUCAGUUGGGAGAGCGCCAGACUCU<sub>crosslinker</sub>APO<sub>4</sub> 3' (SEQ ID NO: 25)

#### **Ligation Of The Two Half-Molecules Of The tRNA Analog**

The above product is treated with T4 polynucleotide kinase in two separate steps to remove the 3' phosphate and add a 5' phosphate.

The newly prepared 5' and 3' half molecules ends are then ligated generally following the previous protocols. The 3' sequences corresponding to the respective 5' sequences are as follows:

Sequence 1B: (Ψ= pseudouridine)

5' PO<sub>4</sub>GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGACUCU<sub>crosslinker</sub>A 3' (SEQ ID NO: 24) corresponded to the 3' half:

5'PO<sub>4</sub>UGGAGGUCCUGUGTΨCGAUCCACAGAAUUCGCACCPur 3' (SEQ ID NO: 31), 3B

and sequence 1B<sub>1</sub>,

5'

OHGGGGCUUUAGCUCAGUUGGGAGAGCGCCAGACUCU<sub>crosslinker</sub> APO<sub>4</sub>

(SEQ ID NO: 25)

corresponded to 3' half

5'PO<sub>4</sub>UGGAGGUCCUGUGTTCGAUCCACAGAAUCUCCACCA3' (SEQ

ID NO: 32).

The latter is recognizable by the aminoacyl tRNA synthetase for alanine in *E. coli*.

The example described above can be used to make and use the SATA, Linking tRNA, and the Nonsense Suppressor tRNA.

**EXAMPLE 4: PLACEMENT OF CROSSLINKERS ON THE mRNA FOR SATA AND NONSENSE SUPPRESSOR tRNA**

In several embodiments, the crosslinker (such as psoralen or a non-psoralen crosslinker) is not placed on the tRNA, but rather located on the mRNA. For example, in one embodiment, the SATA comprises a puromycin located on the tRNA, while the crosslinker is on the mRNA. In yet another embodiment, the Nonsense Suppressor tRNA is used, and this comprises a tRNA with no puromycin, with the crosslinker being on the mRNA. Placement of the crosslinker on the message (the mRNA) can be accomplished as set forth below. The relevant sequence is as follows:

GGGUUAAACUUUAGAAGGAGGUCGCCACCAUG GUU AAA AUG AAA  
AUG AAA AUG AAA AUG U<sub>crosslinker</sub>AG (SEQ ID NO: 26)

For convenience only, and in one embodiment, a message with both Kozak and Shine Dalgarno sequences that has a large number of methionine codons for <sup>35</sup>S labeling is used.

For 4-thiouridine, 5-bromouridine and 5-iodouridine, the message can be purchased fully-made from Dharmacon, Inc. For aryl azides, the method recited in Demeshkina, N, *et al.*, RNA 6:1727-1736, 2000, herein incorporated by reference, can be used.

For 2-thiocytosine, 2-thiouridine, 5-iodocytosine, or 2-chloroadenosine, the modified bases can be purchased as the 5' monophosphate nucleotide from Ambion, Inc. When psoralen is used as the crosslinker, the modified 5' monophosphate nucleotide is made as above.

The modified 5' monophosphate nucleotides are first incorporated into hexamers to facilitate purification. The construction of uridine containing crosslinkers is shown but in several embodiments, the other bases can be incorporated into both stop and pseudo stop codons using similar techniques:

AUG + pUcrosslinker  $\longrightarrow$  AUGUcrosslinker was accomplished using a similar protocol described above, except a preponderance of AUG was used because of the absence of a 3' protection of the pNcrosslinker. The product was purified by anion exchange HPLC from the excess of AUG. Then 5' pAGbiotin 3' was added with T4 RNA ligase. The 3' biotin was simply a convenient 3' blocking group available from Dharmacon. The resulting AUGU<sub>crosslinker</sub>AG<sub>biotin</sub> was again purified followed by 5' phosphorylation and ligated to:

GGGUUAACUUUAGAAGGAGGUCGCCACCAUGGUUAAAAUGAAAAU  
GAAAAUGAAA (sequence M1) (SEQ ID NO: 27)

to produce

GGGUUAACUUUAGAAGGAGGUCGCCACCAUGGNNA<sup>+</sup>AAAAUGAAAAU  
G AAAAUGAAAAUGU<sub>crosslinker</sub>AG<sub>biotin</sub> (SEQ ID NO: 28)

The yield is high enough to obviate purification. Accordingly, using the protocol described above, SATAs and Nonsense Suppressor tRNAs can be made and used in accordance with several embodiments of the present invention.

#### **EXAMPLE 5: USING tRNA SYSTEMS THAT DO NO NEED PUROMYCIN**

Several embodiments of the present invention provide a system and method that do not require puromycin, puromycin analogs, or other amide linkers. In one embodiment, Linking tRNA Analogs and Nonsense Suppressor tRNAs do not require puromycin and can be made and used according to the following example.

For systems without puromycin, a translation system to aminoacylate the tRNA can be used. In other embodiments, aminoacylation can be accomplished chemically. One skilled in the art will understand how to chemically aminoacylate tRNA. Where translation systems are used, any type of translation system for aminoacylation can be employed, such as *in vitro*, *in vivo* and *in situ*. In one embodiment, an *e*-coli translation system is used. An *E. coli* translation system is used for systems with a tRNA modified to be recognized by the aaRS<sup>Ala</sup>. In one embodiment, this is preferable for systems without the stable acceptor (e.g. the puromycin)

3 mcg of each of the following mRNA's are translated in 40 microliters each of Promega S30 E. coli translation mixture:

- a) GGGUUAACUUUAGAAGGAGGUCGCCACCAUG GUU AAA AUG  
AAA AUG AAA AUG AAA AUGUcrosslinkerAGbiotin (SEQ ID NO: 28) and
- b) GGGUUAACUUUAGAAGGAGGUCGCCACCAUG GUU AAA AUG  
AAA AUG AAA AUG AAA AUGUAG (SEQ ID NO: 29)

3 mcg of amber suppressor tRNA manufactured as above are added to the first. 3mcg of suppressor with crosslinker on the anticodon are added to the second. 35S-methionine is added to both and the mixtures are then incubated at 37° C for 30 minutes. The reactions are then rapidly cooled by placement in an ice bath, transferred to a flat Petri dish and floated in an ice bath so that the mixture is 1.5 cm below a ~350 nm light source. They are exposed at ~20 J/cm for 15 min.

After irradiation, the mixtures are phenol extracted and ethanol precipitated. In this manner, systems such as the Linking tRNA Analogs and Nonsense Suppressor tRNAs are aminoacylated and used to connect the message (mRNA) to its coded peptide in accordance with several embodiments of the present invention.

#### **EXAMPLE 6: ALTERNATIVE SEQUENCES**

In a preferred embodiment, Fragments 1, 2 and 3, described above in Example 1, have the following alternate sequences:

Fragment 1 (SEQ ID NO: 13):

5' PO4 GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA N3-Methyl-U 3'

Fragment 2 (SEQ ID NO: 14):

5' UCUAAGΨCΨGGAGG 3'

Fragment 3 —Unchanged from the sequence listed above (SEQ ID NO: 6):

5' PO4 UCCUGUGTΨCGAUCCACAGAAUUCGCACC Puromycin 3'

Using the methods described above, the sequence of alternative Fragments 1+2+3 was (SEQ ID NO: 15):

#### **EXAMPLE 7: APPLICATION TO SARS**

##### **Diagnostic Test for SARS Virus**

In one embodiment a diagnostic test for the SARS virus is provided. The SARS genome sequence is known and the position of associated structural proteins spike (S), membrane (M), nucleocapsid (N) and envelope (E) on the genome are known (Marra,

et al, Scienceexpress/ www..scienceexpress.org / 1 May 2003 / Page 1/ 10.1126/science.1085953 and Rota, et al, Scienceexpress/ www..scienceexpress.org / 1 May 2003 / Page 1/ 10.1126/science.1085952), all herein incorporated by reference).

The "S" protein is associated with binding to target cells and, amongst the coronavirus strains, it is unique to SARS-CoV (Rota, et al, Scienceexpress/ www..scienceexpress.org / 1 May 2003 / Page 1/ 10.1126/science.1085952), herein incorporated by reference. Current R-PCR, EM, and FEIA assays are not adequate because they take too long to perform or have low sensitivity and are thus of limited value in the early stages of disease (Tsi, et al, Emerg. Infect. Dis. 9:9 (2003) and Tsang, et al, Emerg. Infect. Dis. 9:11 (2003)), all herein incorporated by reference. The virus is readily available in the sputum (Hsueh, et al, Emerg. Infect. Dis. 9:9 (2003)), herein incorporated by reference. The "S" protein is present in all of the SARS-CoV strains (tor2, Urbani, TW-1, HKU-39849, and CUHK-W1).

One diagnostic test available today is the Nanoparticle-Based Bio-Bar Codes technology (Nam, et al, Science 301:1884-1886 (2003)), herein incorporated by reference. This method appears to have extreme sensitivity, which should enable one to detect the low levels of viral particles found in sputum in the early stages of SARS disease. There are other, faster to perform methods which can occur in real time, however they are less sensitive in some cases. Several embodiments of the present invention facilitate the development of the reagents needed for the assay in a matter of several days instead of weeks or months.

After production of adequate amounts of pure "S" protein, several embodiments of the present invention can be used to make at least two additional reagents: two highly specific binding proteins that bind to two different protein domains on the "S" protein, one for use as the trapping probe and the other for the signaling probe.

In one embodiment, the protocol will be performed as follows:

#### *Preparation of Test Reagents*

In one embodiment, the reagents will be prepared in the following manner:

##### **A. Preparation of purified "S" protein**

1. The SARS-CoV genome sequence will be obtained from Genebank (Accession #AY274119-3) from which the portion of the sequence that codes for "S" protein will be obtained.

Primers for cDNA of  
 5'PO4GCGGAUUUAGCUCAGUUGGGAGAGCGCCAGA(N3-  
 MethylU)UCUPsoraleAAGΨCΨGGAGGUCCUGUGTYCGAUCCACAGAA  
 UUCGPuromycin 3' (SEQ ID NO: 33)

For Linking tRNA Analog and Nonsense Suppressor tRNA, the above sequences are similar, except adenosine is used to replace puromycin.

While a number of preferred embodiments of the current invention and variations thereof have been described in detail, other modifications and methods of use will be readily apparent to those of skill in the art. For all of the embodiments described above, the steps of the methods need not be performed sequentially. Accordingly, it should be understood that various applications, modifications and substitutions may be made without departing from the spirit of the invention or the scope of the claims.

#### **FURTHER APPLICATIONS**

Until now, deciphering the mRNA sequence for a protein has been a huge bottleneck for the proteomics industry because it involves a very significant investment in time, effort, and dollars to perform an N-terminus analysis from which the best guess for the mRNA sequence for the protein is made. An N-terminus analysis involves chemical dissociation of the protein so that its amino acids and their order in the protein are determined. In one embodiment, the current invention solves this bottleneck problem by linking the exact mRNA message with its cognate protein during the translation process, thereby providing the user with the ready blueprint for making more of that protein, and obviating the need for N-terminus analysis. For example, one method for using binding proteins as probes for "S" protein can be performed as follows:

1. The initial mRNA library will be created by codon iteration.
  - a) A proprietary method of constructing a set of messages from random codons that do not include a stop codon will be used to create a reading frame.
  - b) An appropriate stop codon will be added. 3' and 5' untranslated regions will also be added to each oligo.

- c) This will create on the order of  $>10^{14}$  different messages, each averaging 128 codons or more.
2. When a protein is needed *in vivo*, the genes that code for that particular protein are activated to generate an mRNA sequence code for that protein. The mRNA then carries the message from the nucleus to the cell's ribosomes, where amino acids are assembled into the designated protein in the order dictated by the mRNA code. Upon completion, the protein and the mRNA are enzymatically dissociated from the ribosome and from each other. If one is interested in the "S" protein sequence will be made.
3. A cDNA copy of and wishes to produce more of it, or wishes to modify the sequence will be made.
4. PCR will be used to prepare enough cDNA to insert into *E. coli*.

The newly generated "S" protein will be harvested from *E. coli* using established methods (Doonan, ed., Vol. 59, New Jersey: Humana Press (1996), herein incorporated by reference to enhance desirable properties or to eliminate undesirable ones, the protein's mRNA code must first be known.

Further embodiments of the present invention comprise the "Linker System", a revolutionary system of novel compounds and methods which enable scientists to quickly and easily chemically link proteins to the mRNAs that encode them (see Figure 14).

#### **Novel Proteins**

In one embodiment novel proteins for specific purposes can be made quickly by translating embodiments of the mRNA library *in vitro* as shown in Figure 14 (B) above using embodiments of the linking system. The resultant library of protein-Linker-mRNA complexes will number in the trillions ( $10^{14}$ ) of protein variants from which the ones with the desired properties can be selected. In one embodiment the mRNA from the selected protein-Linker-mRNA complexes can then be chemically cleaved off of the complex and used to produce large amounts of the protein either in Bioreactor cultures, or by large scale *in vitro* translation. In another embodiment, if better versions of the protein are desired, the mRNA can then be subjected to established accelerated protein evolution techniques, and the resultant mRNA library can then be translated again in the Linker System to rapidly evolve huge libraries of protein variants. The



proteins with enhanced characteristics may then be selected from the rest. In another embodiment the process can then be repeated multiple times to strengthen the desired trait or to add additional traits to the protein, since all of the mRNAs to the best candidate proteins are attached to their respective proteins and, therefore, can be easily harvested for repeat cycles (Figure 15). Preferred embodiments, thereby, preferably obviate costly, time consuming conventional procedures currently used by the industry to first identify the exact mRNA sequences that code for proteins of interest.

#### **Native Proteins**

In one embodiment native proteins can also be linked with their cognate mRNAs and easily selected in the same way as for novel proteins, described above. In one embodiment, mRNAs collected from any organism or of unknown origin are translated *in vitro* using the SATA reagent and linking system as shown in Figure 14 (A) above. The proteins of interest are then selected out of the resultant library of protein-Linker-mRNA complexes.

In one embodiment, this approach is particularly useful for identifying the genes responsible for specific proteins because the mRNA message for a particular protein reflects the genes that created the mRNA code for that protein. Therefore, the mRNA message can be used as a probe to go back and identify the exact gene or gene sequences and their location in the genome that encode for that particular mRNA, and therefore for that specific protein.

In another embodiment adapted to microarrays, this allows one to quickly derive gene activity profiles characteristic for specific disease states. Gene activity profiles have already been used successfully to establish accurate diagnosis for specific types of cancer. Several embodiments of the invention enable one to establish such profiles much faster and more efficiently than with the technologies currently used in the industry. Further, since preferred embodiments identify both the protein products and the genes that code for them, these embodiments preferably can be used to rapidly evolve proteins that target either the protein product associated with the disease or the genes that are expressing the disease-related protein.

The advantages of some embodiments of the invention are summarized below.

*Ability to speed new protein development:* In one embodiment, the ability to link proteins with their mRNAs greatly simplifies the development of protein based

products because it obviates the time, cost, and effort intensive N-terminus analysis method currently used to determine the exact mRNA sequence for each protein. Once identified, utilizing several embodiments of the invention, the mRNA can be used to quickly generate more of the protein, or can be modified by introducing mutations *in vitro* to produce desired variants of the protein. In preferred embodiments, novel proteins can therefore be produced in weeks, rather than months or years required by the current method.

*Ability to optimize new protein development:* In one embodiment, the invention enables one to rapidly optimize development of proteins with desired properties by creating huge mRNA libraries from which rare proteins can be selected that are linked with their mRNAs, using *in vitro* translation.

*Ability to greatly reduce manufacturing costs:* In one embodiment, manufacturing of protein based human therapeutics and vaccines using *in vitro* translation procedures and translation formulations do not include reagents derived from animals and, therefore, greatly lowers the overall manufacturing costs. Manufacturing human therapeutics currently requires animal products such as blood serum for production. This is expensive and creates additional costs to assure that the animal products used do not contaminate the therapeutic products with prions, viruses, or other animal borne contaminants.

The applications for some embodiments of the invention extend to virtually every area in which proteins are involved. The following areas provide additional non-limiting examples of these applications, and potential products.

#### **Human Therapeutics**

In one embodiment, binding proteins can be made quickly and easily that bind any reliable cell surface marker, including those found on diseased cells such as malignant cells, which when bound induces cell death. In another embodiment, binding proteins can also be made to key biochemical targets such as Macrophage Migration Inhibitory Factor that, when inactivated by protein binding, prevents onset of Type I diabetes. Additionally, preferred binding proteins can be used as therapeutic cell growth factors, for example ulcerative colitis can be effectively treated with Epidermal Growth Factor like proteins which stimulate re-growth and healing of the gut epithelium.

In one embodiment, binding proteins can also easily be made that bind established surface markers on disease causing organisms, thereby effectively inactivating them. Additionally, preferred binding proteins can be used in diagnostic tests for detecting these organisms. Targets for such binding proteins include, but are not limited to, the viruses that cause HIV, Hepatitis, Herpes, smallpox, West Nile virus, SARS, viral pneumonia, genital warts and any other well characterized virus. Also, preferred binding proteins can target the bacteria that cause Anthrax, bubonic plague, botulism, drug resistant staphylococcus, cholera, bacterial pneumonia and any other bacteria. Fungi such as pathogenic yeast can also be inactivated by successfully targeting them with preferred binding proteins. In another embodiment, proteins can also be easily selected that block the binding sites on the host's cells which the infecting organisms target, or that their toxic metabolic products target. As such, the host is protected from the pathogen and its toxins.

#### **Diagnostics**

One embodiment of the invention is ideally suited for making highly accurate and stable diagnostic tests. Preferred embodiments can be used to identify the best target reflective of a disease state or any condition of interest, and subsequently, to generate novel binding proteins for use as trapping and/or signaling reagents. Preferred binding proteins can be selected with sensitivity, specificity and/or stability properties that are superior to monoclonal antibodies without the cost, time and effort associated with producing monoclonal antibodies for this purpose.

In one embodiment, preferred proteins bind to the well characterized cancer targets CD-22 for fluid cancers or CD-33 for solid tumors. Preferred binding proteins may also be created that are substantially free of the immunogenicity and/or manufacturing problems associated with mAbs to these antigens, yet preferably still retain the same or better binding, specificity and/or sensitivity properties as the commercially available mAb products on the market.

#### **Food and Drug Administration (FDA)**

Overall, the FDA is trying to expedite the approval process for mAb based healthcare products. They find, however, that there are problems with mAbs such as: the inadequate characterization of the mAb and its target, changes in mAbs during scale up production, and immunogenicity of the mAb. Immunogenic reactions, while lower

than with murine mAbs, still remain around 8% for the chimaric mAbs. Also, plant glycans from mAbs raised in plants can cause immune reactions, and animal serum and other animal products used in tissue culture are an especially serious concern for the FDA because of contamination concerns with mad cow disease (prions), and hoof and mouth disease, etc. The result is that much expensive monitoring to document safety is required by the FDA. This is a particularly serious problem for European companies trying to obtain FDA approval for their mAb products because of the mad cow "prion" problem associated with European cattle. In one embodiment, the solution to this problem is to use preferred methods to generate production sized amounts of the preferred binding protein with an *in vitro* translation system that uses synthetically formulated translation mixtures that do not involve animal products. Because of this, the FDA has indicated that the approval process for such antibody substitutes will likely be faster than for mAb products (personal communications, April 14, 2003). As such, in one embodiment, binding proteins would only need to be safe and efficacious when compared to approved mAbs for the same targets.

By producing binding protein embodiments with equivalent therapeutic value but without the manufacturing expenses, high costs, difficult FDA hurdles, and side effect problems associated with mAbs, preferred embodiments of the mAb substitute products may receive strong interest from current mAb users and manufacturers.

In another embodiment, surface plasmon resonance technology may be used in combination with preferred methods for isolating and enriching rare proteins out of mRNA libraries which exhibit chosen properties.

In one embodiment artificial translation mixtures are used to replace currently used animal reticulocyte based translation mixtures. Preferred embodiments may be adapted to large scale translation systems for production of large amounts of preferred protein products.

In another embodiment CHO cells may be cultured in a Bioreactor. Preferably the mRNA for the selected proteins will be incorporated into the genome of the CHO cells. In another embodiment, the CHO cells grown in the Bioreactor culture will be selected that express the protein coded for by the inserted mRNA. In a further embodiment, the preferred target protein may be isolated from the CHO cells or culture medium and further purified.

In another embodiment, preferred methods produce an initial binding protein that binds to well characterized cancer targets such as CD-22 or CD-33. Proteins may be selected that preferably do not have the same negative side effects associated with currently marketed monoclonal antibody products that target these binding sites.

#### **Cellulase Enzymes**

In another embodiment, preferred methods produce an enzyme that substantially breaks down cellulose to glucose. Food and beverage producers convert edible corn starch, from corn kernels, to glucose with the enzyme amylase. Glucose is used as a sweetener in food and soft drinks and is used in the fermentation process to make alcoholic beverages or ethyl alcohol for use as a gasoline additive.

The non-edible part of the plant is composed primarily of cellulose and is currently not used for glucose production. Chemically, cellulose is a long chain of glucose molecules. Therefore, in one embodiment cellulase enzymes that digest the cellulose part of the plant to glucose would allow one to use substantially the entire plant for the production of glucose, instead of just the corn starch component. With such preferred enzymes, considerably more glucose could be produced from the same amount of biomass. Further, with these preferred enzymes, virtually any plant material could be used to make glucose. This would translate into more cost effective end products and, therefore, these preferred enzymes should be of great interest to food and alcohol producers.

#### **State of the Art**

Cellulase is currently produced for research purposes by the Danish firm Novozyme Corporation. They isolate the enzyme from two microbes, *Aspergillus niger* and *Trichoderma reesei*, in a bioreactor process called submerged fermentation. Novozyme has attempted, but has not been successful, in isolating a cellulase from these organisms that effectively breaks down non-edible parts of plants on a large scale. In preferred embodiments, the invention may be used to rapidly create a family of cellulases that will digest any cellulose to glucose. In another embodiment, the gene sequences for preferred enzymes can then be inserted into any convenient organism for large scale production.

#### **mRNA Libraries**

In one embodiment, the invention can be carried out completely *in vitro* and may also provide huge mRNA libraries. Preferred embodiments use a linker which acts as a protein acceptor and the linkage takes place only after the protein has been completely synthesized. Therefore, in preferred embodiments the correct mRNA message is attached to the right protein, and because synthesis stops only after the entire mRNA message is translated, it is virtually impossible to end up with shorter than normal proteins or with the message on the wrong protein. In one embodiment, the invention preferably does not suffer from the same problems that limit the utility of competitor's technology and therefore, has much greater and wider-ranging application. Preferred embodiments are a much more powerful technology for creating and selecting proteins for commercialization, in addition to its potential use in microarrays.

In one embodiment, a binding protein approach provides significant advantages over traditional mAbs. In another embodiment, the invention can make vaccines safer and/or more effective, which would preferably result in less exposure to product liability. In another embodiment, preferred methods produce mRNA libraries for use with, for example, companies that use microarrays to establish serum protein profiles that reflect disease states such as cancer, or to identify gene or biochemical targets for therapeutic intervention. Also included, but not limited to, are companies that tap commercially available mRNA libraries for mRNAs that yield binding proteins with therapeutic value. Also included, but not limited to, are major companies which produce diagnostic tests. Preferred embodiments of the invention can also be used to create binding proteins that preferably result in diagnostic assays with higher sensitivity, specificity and/or accuracy for various items, including, but not limited to, cancer markers, for infectious diseases such as hepatitis, AIDS, SARS, H. pylori, and genital herpes, as well as for other disorders such as colitis and autoimmune disorders.

Preferred biowarfare embodiments provide opportunities covering a wide-ranging spectrum of firms that range from start-ups to large established companies, as well as the Federal Government. These institutions are developers of diagnostic tests, anti-toxin therapeutics, neutralizing agents and vaccines which can be used with preferred embodiments.

In the agriculture field, preferred embodiments include, but are not limited to, animal therapeutics and diagnostics, and treatment for plant pathogens.

Industrial users of preferred embodiments of the invention include, but are not limited to, companies that design and/or produce enzymes for use in industry, companies that are following the current trend of adapting enzymes to reduce production costs in the food and petroleum additive business, and the paper, lumber and petroleum industries for managing and controlling their environmental waste.

#### **EXAMPLE 8 - APPROACH IN DIAGNOSTICS**

##### **Target Identification**

In the broadest sense, preferred embodiments could be used to identify a target that is over expressed in a single patient or more broadly by all or most patients with a specific disease. Preferred embodiments preferably take advantage of the ability to link mRNAs with the proteins they code for. For example, to identify a target, all mRNAs isolated from the serum of a patient or patients with a specific cancer, can be translated in vitro using standard eukaryotic or prokaryotic in vitro translation systems plus the SATA linker system. A protein profile of the resultant mRNA-SATA-Protein complexes can identify proteins that are over expressed as compared to normal patients. Establishing such protein profiles is a well established technique used in therapeutic proteomics today. One advantage of the approach of preferred embodiments is that the mRNAs are attached to the proteins and therefore, they can be harvested off of the selected proteins for further development of an assay. Scaled up amounts of the selected mRNAs can be made by reverse transcription of the mRNAs and PCR of the resultant cDNAs. The cDNA can be transfected into a host organism (e.g., E. coli, yeast, CHO cells, etc.) from which the proteins would be harvested. These proteins are the targets from which the one that best identifies the disease is chosen empirically and used for standards in the assay.

##### **Trapping and signaling binding proteins**

In one embodiment, the task is to identify two binding proteins that are highly specific for two separate binding sites on the target protein (T) and not on other serum proteins and that are also highly stable under test system conditions.

The following protocol illustrates several embodiments of how these binding proteins can be produced:

1. An initial mRNA library can be constructed in one of two ways.

##### **A. First Method**

*Initiation sequence*

- An RNA oligo that includes a 5'UTR region leading to an AUG start codon can be constructed by commercial means. This would be used as a reagent that is later ligated to an mRNA library constructed from a random assembly of RNA triplets. This oligo sequence is necessary to initiate translation.

*Random mRNA library*

- A series of up to 61 RNA triplets that make up all of the sense codons can be commercially synthesized for the company. The synthesis would include OH groups on both the 5' and 3' ends of all of the triplets.
- All of the oligos would be highly purified to exclude potential reading frame shifts.
- A portion of these triplet oligos would be 3' protected with any of the commercially available 3' protective groups.
- A portion of the protected oligos would be ligated in a random fashion to a portion of the unprotected oligos with T4 RNA ligase.
  - This first ligation would form an oligo that includes the first two-codon sequences. Then one half of this material would be 5'phosphorylated and the other half would be 3' deprotected. The two pools would then be ligated to each other with T4 RNA ligase as before.
  - This second ligation forms an oligo that includes 4 codon sequences. Repeating this procedure 7X results in an mRNA library that includes up to 128 random codons.
  - The randomized 128 codon oligo can then be ligated onto the 5' UTR start sequence. A stop codon and a 3' UTR can be attached at this time as well.

**B. Second Method***Random DNA library*

- This method involves using highly purified phosphoramidite trimers to construct a randomized DNA library.
- Available trimers are: AAA, AAC, ACT, ATC, ATG, CAG, CAT, CCG, CGT, CTG, GAA, GAC, GCT, GGT, GTT, TAC, TCT, TGC, TGG, TTC.
- Highly purified phosphoramidite trimers can be purchased from a company such as Glen Research Corporation, Sterling, Virginia.
- The randomized library can be constructed using the same principles as described above, and the reading frames can be inserted between 5' and 3' UTR's as above as well, using T4 DNA ligase

**2. Linking mRNA with its cognate protein with the SATA linker.**



- a) The library will be translated *in vitro* using commercially available prokaryotic or eukaryotic translation systems.
  - b) The mRNAs will be connected to their peptide sequences using the SATA linker technology and UV irradiation at about 320-400 nm system.
3. To determine the affinity constant distribution of the proteins coded for by the library, the SARS "S" protein produced above will be attached to avidin coated membranes through a biotin linker attached to the "S" protein, or by an anti-"S" antibody attached to the membranes, or by some other convenient means.

The protein-SATA-mRNA complexes from the random library will be reacted with "S" protein coated surface plasmon resonance (SPR) membranes. The affinity constant distribution of the set of proteins will be established by titrating varying quantities of stationary target against the protein population as shown in Fig. 16. This will generate the distribution of binding constants for the protein library-1.

To evolve proteins with higher binding constants, the above distribution will be used to calculate the total amount of "ST" protein needed to select the proteins with the highest affinity required for use as trapping ( $P_{\text{trap}}$ ) and signaling ( $P_{\text{sig}}$ ) probes for the assay as well as the number of rounds of selection necessary to attain the required affinity (see Appendix 2 for more detail). The amount of "ST" protein determined by the above will be bound to membranes a stationary phase as before and will be allowed to react with the protein-SATA-mRNA library. The resultant "ST"-protein-SATA-mRNA complexes will be recovered and irradiated briefly with 313 nm light to disassociate the mRNAs from the complexes. The mRNAs will then be reverse transcribed and amplified with error prone PCR. This process will be repeated until proteins with optimal binding properties are evolved.

Preparation of the  $p_{\text{trap}}$  and the  $p_{\text{sig}}$  probes. The Intrinsic sensitivity will depend on a high affinity for "ST" protein and specificity will depend on a low affinity for proteins other than "ST" protein.

- a) The mRNAs from the selected high affinity binding proteins will be reverse transcribed and amplified with PCR and will be inserted into *E. coli* for large scale production of each protein. The proteins will be

harvested using established methods (Doonan, ed., Vol. 59, New Jersey: Humana Press (1996)), herein incorporated by reference.

- b) The proteins will be tested for cross reactivity with "ST" proteins from other corona virus strains, sera from normal patients, benign diseases of the same tissue, etc. using SPR. The trapping and signaling proteins for the assay will be chosen from the population of proteins that do not substantially cross react.
- c) "ST" or "T" protein on SPR membranes will be reacted with one of the high affinity proteins (P1) at a concentration that saturates all of the binding sites on the "ST" protein for P1. The reacted membrane will then be reacted sequentially with titrations of the other proteins (P2, P3, Px) until one is identified that also demonstrates maximal binding. This protein will thereby bind to a domain on the "ST" protein other than the P1 domain. These proteins will become the  $P_{\text{trap}}$  and  $P_{\text{sig}}$  probes as shown in Figure 17.
- d) The  $p_{\text{trap}}$  probe will be functionalized by applying the trapping protein to 1  $\mu\text{m}$  polyamine micro particles that have magnetic iron oxide cores (Nam, et al, Science 301:1884-1886 (2003), herein incorporated by reference, e.g., Figure 18). In one embodiment, the reagents can then be adapted to any test format and testing device that is formatted to use all or any combination of the reagents.
- e) The  $p_{\text{sig}}$  probe will be functionalized by applying the signaling protein and bar code oligonucleotides to 30 nm gold beads as shown in Figure 19.

#### *The Assay*

A. In one embodiment, the assay will involve the following steps:

- The trapping probe, the patient sample, and the signaling probe will all be reacted together in a reaction well.
- If the SARS virus is present, a complex will form consisting of the SARS virus sandwiched between the trapping and signaling probes via the exposed "S" protein.

- The complex will be isolated magnetically from the non-binding components. The non-binding components will be washed away with 0.1 M phosphate buffered saline.
  - The hybridized DNA oligonucleotides will be dehybridized with NANOpure water to release the single stranded DNA bar code sequences.
  - The concentration of the bar code sequences will be measured with a Scanometric DNA reader.
  - Viral titer will be quantified by extrapolation from an “S” protein standard curve.
  - The assay will be adapted to large scale testing using automated testing devices designed around magnetic beads.
  - This test system has been used by others to detect PSA at concentrations as low as 300fM to 3aM in a 10 $\mu$ l test volume. The extreme sensitivity and specificity associated with this assay will make it possible to detect the low virus levels found in sputum of SARS infected patients in the early stages of disease.
- B. The assay's reactions are shown graphically in Figure 20
- C. Test to determine if diagnostic is successful.
1. The SATA based assay will be compared with the R-PCR assay on the same samples to establish added value and efficacy.
  2. The test will be used on sputum from patients with common respiratory virus infections such as influenza type A and B, adenovirus respiratory syncytial virus and parainfluenza virus types 1, 2, and 3 to determine specificity in a clinical setting.
  3. The assay will then be used on sputum from SARS patients to establish sensitivity, specificity, accuracy values for the assay, and to determine its usefulness for early detection and management in a “real world” clinical setting.

#### *Vaccine for SARS Virus*

In one embodiment, a vaccine to the SARS virus is provided. Methods for producing vaccines described herein are prophetic. In one embodiment, rather than select for a few proteins with the highest binding affinities in a given distribution, one

can use a less stringent selection so as to have a high number of different sequences and use multiple rounds of mutation with gradual increase in the stringency to evolve a large population of proteins with a high binding affinity. Such proteins are of value for making vaccines. The logic is similar to an anti idiotypic vaccine except that there will be one and only one surface epitope that can react with the immune system. The aggregate concentration of the "S" protein presented to the immune system by the family of proteins will be sufficiently high to reach the threshold level required to stimulate a T-cell and B-cell response. However, the concentration of any single protein within the family will be below the threshold required to stimulate a response to that protein. Therefore, the vaccine will stimulate antibody production only against the "S" epitope and not against any of the other epitopes present on the family of proteins. This will prevent production of antibodies that could inactivate the vaccine. In another embodiment, the vaccine will be synthesized such that it will stimulate antibody production against the "S" epitope and one or more other epitopes that have either a neutral or synergistic effect with activation of the "S" epitope.

In one embodiment, one method for making a vaccine to the SARS virus using such de-selected proteins can be performed as follows:

#### 1. Preparation

- a) The sequence for the major histocompatibility complex class II (MHC-II) will be added to all of the cDNAs of a random mRNA library. The MHC II-binding sequence will permit the appropriate T-cell and ultimately, B cell response.
- b) The library will be transcribed and translated *in vitro* using prokaryotic or eukaryotic translation systems and the SATA linker to link the proteins to their cognate mRNAs.
- c) The proteins will be selected from the library with a probe specific for the "S" epitope. The probe will be chemically linked to SPR membranes.
  - i. if the probe is an antibody, antibodies will be used with random idiotypes as the blank for deselection as shown in Figure 21.

- ii. if the probe is another type of aptamer, the aptamer will be saturated with S.
- d) Proteins that have high affinity for the anti-S antibody will be selected. The mRNAs will be dissociated from the complexes with 313 nm UV light and will be reverse transcribed and mutations will be introduced into the cDNA using error prone PCR. The mutated cDNA library will be transcribed and translated as described in 1.b above and the proteins with highest affinity will be selected as before. This process will be repeated until a population of proteins is achieved that has maximum affinity for the anti S antibody.
- e) The mRNAs will be dissociated from the complexes with 313 nm UV light. Then the mRNAs will be reverse transcribed for each mRNA.
- f) PCR will be used to make enough cRNA to insert the sequence into a vector to insert into either *E. coli* or yeast for bioreactor culture or alternatively for direct translation in a scale-up in vitro translation system or for insertion into animal genomes such as goats for selection of product out of the goat milk.

## 2. Expected Immune response to the protein library vaccine

The scheme shown in Fig. 22 illustrates B-cells with "S" receptors that become activated by the MHC-II on the proteins and in one embodiment, production sized lots of each of the reagents can be produced in bioreactor culture or in any host organism of choice. Because of the rapidity of the reactions involved, preferred reagents can be produced in weeks rather than many months or years and at considerably lower costs than required for making hybridomas for monoclonal antibodies. Although the above example discusses SARS, one of skill in the art will understand that other vaccines can be prepared according to the methods described above.

In another embodiment, the task of developing an assay when the target is already known, such as any of the currently used tumor markers, would be expected to be much easier and faster since the task would only involve selection of the best binding proteins for use as trapping and signaling reagents.

### Selection Strategies for the Breeding of Large versus Small Populations of Proteins with Desired Binding Characteristics

In one embodiment, in vitro protein evolution is used to develop novel proteins. In several embodiments, the methods used are based on selection by binding affinity. In some embodiments, this depends on the competition between protein molecules for binding to a target ligand: the proteins that are bound most tightly are selected and reproduced. The stringency of this in vitro selection is determined by the concentration of the unbound target ligand:

$$\frac{[A_i T]}{[A_i]_{tot}} = \frac{[T]}{(k_i + [T])}$$

Where T is the target ligand,  $k_i$  is the dissociation constant of  $A_i$ , any given protein molecule  $i$ ,  $[A_i T]$  is the concentration of that protein that is bound to the target and  $[A_i]_{tot}$  is the sum of the concentration of the bound and the unbound or total protein  $A_i$ . This makes

$$\frac{[A_i T]}{[A_i]_{tot}}$$

the fraction of the total of the total protein with binding constant  $k_i$  that is bound. Likewise, the amount of enrichment by binding to ligand between any two proteins depends on the  $[T]$  and the ratio of their affinities:

$$\frac{(k_j + [T])}{(k_i + [T])}$$

Notice that the ratio of the bound proteins differs from the ratio of the starting or total protein by the factor \*. This can be called the enrichment factor because it is the amount of enrichment that can be achieved by binding to the target. The value of this factor is determined by the relation between the concentration of the free target ligand  $[T]$  and the  $k$  values. If the  $k$  values are much smaller than  $[T]$  the factor is 1 and there is no enrichment. If the  $k$  values are much bigger than  $[T]$  the enrichment is maximized at the ratio of the  $k$  values.

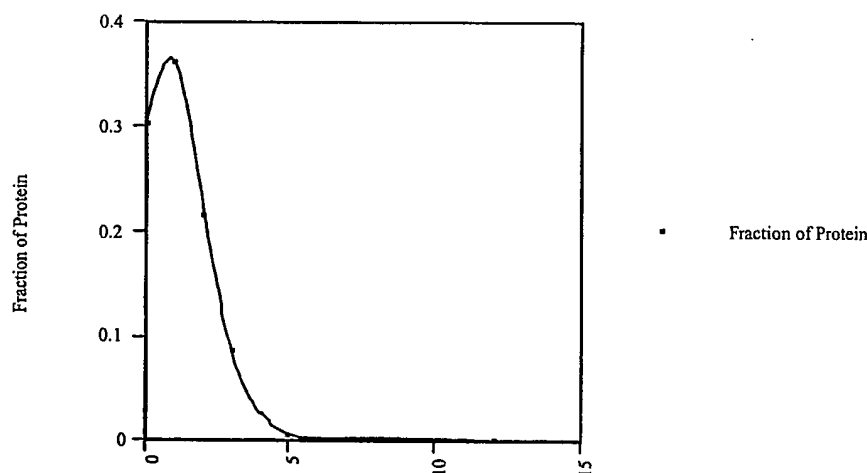
Note that there are two ways to control the magnitude of the  $[T]$ . One is to use a sufficient volume that the total protein concentration is small enough to not effect the  $[T]$  even if all of the protein is bound:

$$[T]_{tot} = [T] + \sum [A_i T]$$

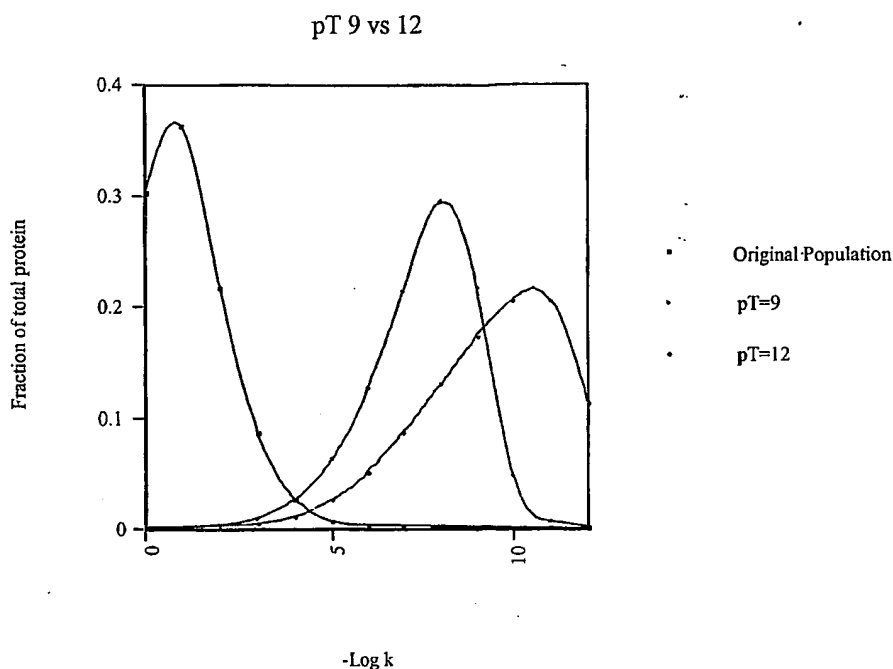
If the total T is sufficiently larger than the protein concentration than it is essentially equal to the unbound T. The other is to know the distribution of the k values of the protein, postulate a desired  $[T]$  value, and use eqn. (1) above to find the necessary  $[T]_{tot}$ . The selection of the  $[T]$  value to use will depend on the end result desired. For some purposes a very stringent selection (low  $[T]$  value), with or without mutation, may be most useful. This would find a rare number of proteins when a small number of tightly binding proteins are desired. The other extreme would be to use much less stringent selection coupled to mutation in an iterative fashion. This would breed a whole population of high affinity proteins. Of course, in practice a combination would probably be used.

Imagine a hypothetical k distribution consistant with Lancet et al.:

Now consider the effect of using a  $[T]$  value of  $10^{-9}$  vs  $10^{-12}$  to harvest bound



protein connected to their cognate mRNA's and then reverse PCR the mRNA's to obtain a concentration equal to the original and transcribe and translate:



If the starting population has  $10^{11}$  different proteins the  $[T]=10^{-9}$  selection will recover just under 1 million of these, whereas the  $[T]=10^{-12}$  selection will only recover a few hundred. For a target with less affinity this would be even less. Mutating the red curve and subsequently harvesting at pT values of slightly greater value will generate a very large number of proteins with steadily higher binding affinities.

When might this slower strategy be preferable? If one wants to generate proteins with more than one selection criterion would be one example. For instance, to generate an enzyme one would want to select a protein that binds a transition state analog of the reaction to be catalyzed tightly but bind the substrate much less tightly (Voet and Voet). Arranging the steps so that each criterion in turn had a large number of molecules to select from could facilitate this.

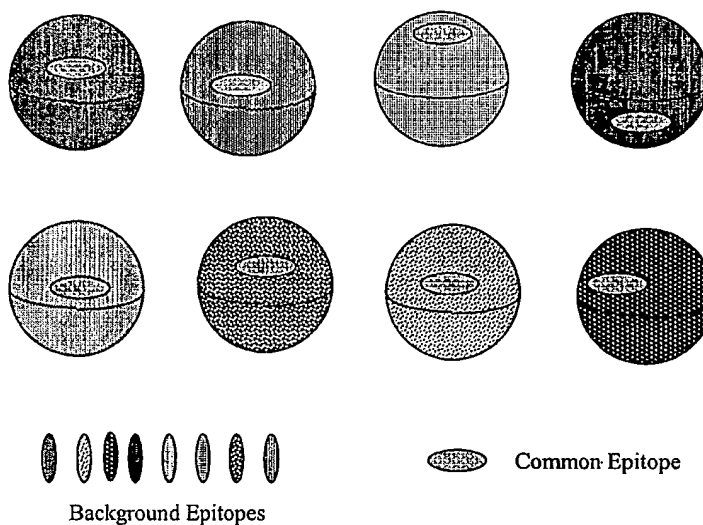
Yet another way to use this is to generate a large number of proteins that have one, and only one, surface characteristic in common.



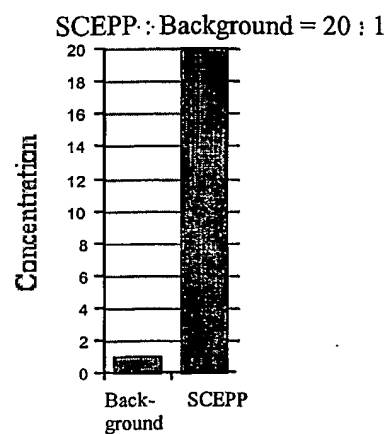
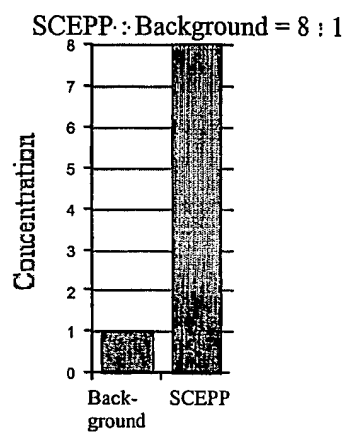
## Use of Single-Common-Epitope Protein Populations (SCEPP) to Interact with the Immune System

Consider such a population below:

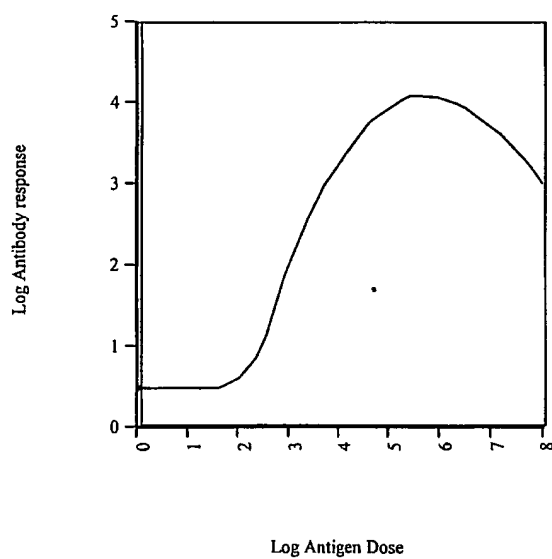
These are proteins that share a common surface binding characteristic. The common surface epitope is the only thing they have in common.



If the population contains these eight proteins the common epitope will be present in eight times the concentration of any of the background epitopes.



This is significant because the induction of a primary antibody response and the subsequent secondary responses are concentration dependent:



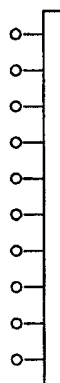
Not only will there be no other common surface epitope, there will not be a common peptide fragment that can be used for T cell recognition. This will further prevent immunization against the one common surface epitope. This means that the SCEPP can be used directly as blocking proteins without the development of serum-sickness-like reactions:

#### EXAMPLE

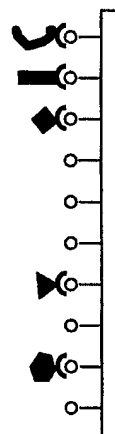
Consider a given Toxin molecule:



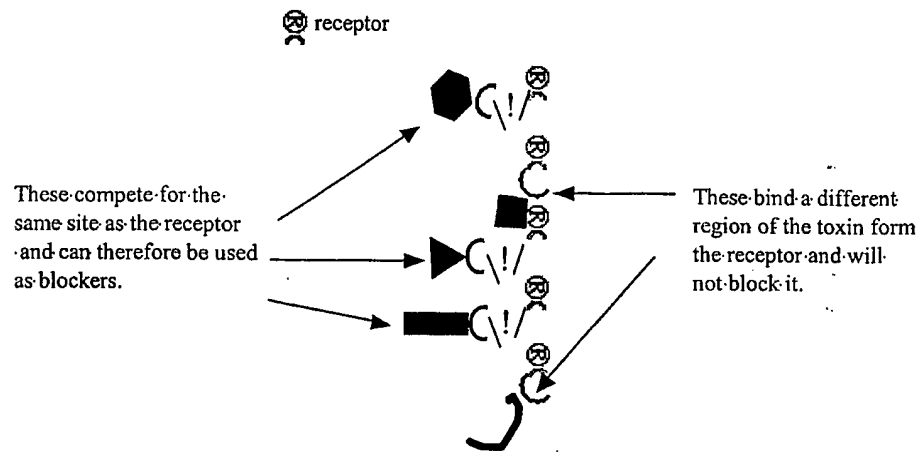
Bind it to a substrate surface, or attach it to a binder such as biotin:



Develop a protein library connected to its mRNA and breed a population of proteins complementary to the toxin.

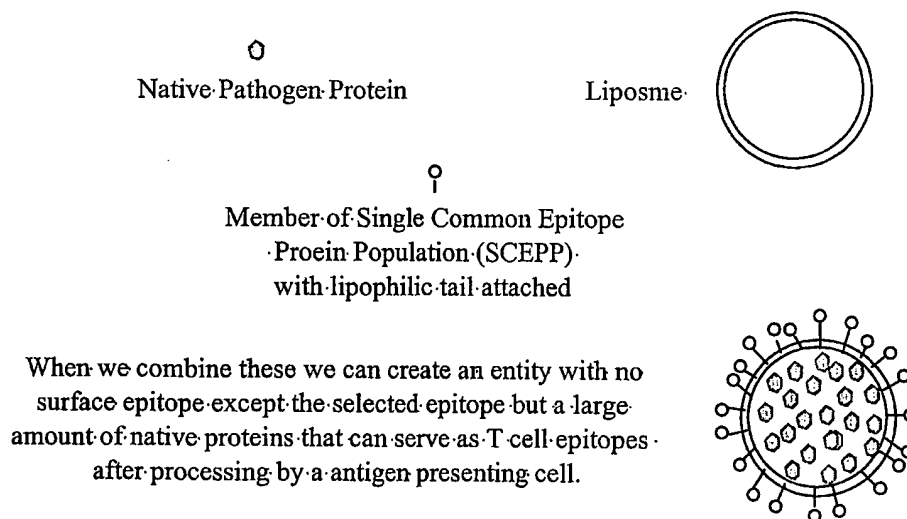


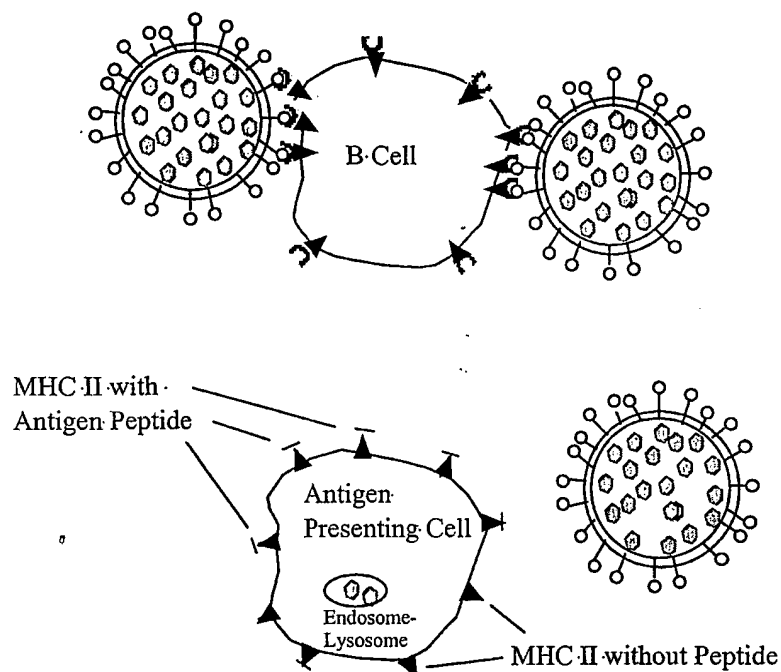
If necessary, expose the bound toxin to its physiological receptor and and re-expose the SCEPP. Discard the bound SCEPP and harvest the unbound. This will select out for proteins that do not block the receptor. Clearly, other forms of competition can be used. This is an example of multiple criteria for binding mentioned above



This method can be used to block toxins, receptors directly, hormones, or viroids without generating an immune response.

Combining a SCEPP with an MHC II binding peptide that does not change the surface epitope could be used to create a vaccine targeted to one and only one surface epitope. As an example consider the following:





If the SCEPP is bred to be complementary to the antigen binding region of a protective antibody from another organism, the result would be a vaccine that reproduced a similar antibody in the vaccinated organism. The logic of this is much like that of an anti-idiotypic vaccine. If the SCEPP were bred to complement a cellular viral receptor the produced antibody would resemble the receptor. In some embodiments, one advantage gained by this technique would be that it allows the selection of the *particular epitope on the protein* that the antibody will bind to. It does not require a native protein within the liposome but using one permits affinity maturation to take place. There are other ways of having an MHC binding peptide, such as using sequences that will remain within the interior of the protein but on digestion by the acid proteases of the APC will be freed to bind to the MHC II.

#### **Biowarfare**

In one embodiment, binding proteins can also be made that neutralize toxins used as biowarfare agents, both those produced by infectious organisms or as weaponized biowarfare agents. Such agents may include, but are not limited to, botulinum toxin, ricin, and anthrax toxin. In another embodiment, Enzymatic proteins

can also be created that neutralize nerve agents which include, but are not limited to, Soman, VX, and Serin. Preferred neutralizing proteins can be incorporated into articles such as clothing to inactivate the agents before they contact the body, and can be used as aerosols to inactivate the agents on surfaces and/or while still airborne.

#### **Agriculture**

In another embodiment, as in human therapeutics, proteins can be created that bind to established surface markers on organisms that cause disease in food crops and livestock. Preferred embodiments can be used, for example, to make inexpensive diagnostic tests, therapeutic treatments and vaccines to such diseases including, but not limited to, anthrax, hoof and mouth disease and mad cow disease for cattle, and Newcastle disease in poultry. Further, in another embodiment for highly infectious organisms such as the one which causes hoof and mouth disease, preferred protein products can be created inexpensively for use in broad spectrum applications that will decontaminate various surfaces, such as barns, corrals, food troughs, feed lots and transportation equipment, thereby preferably preventing the onset of the disease or its spread to other animals or sites.

#### **Industrial**

In another embodiment, enzymatic proteins can also be created that provide maximum production efficiency by preferably functioning at temperature, pressure and/or chemical conditions that are optimum for specific industrial reactions. Currently industry is often forced to use enzymes that function at less than optimum production conditions for lack of better enzyme choices. Novozyme, for example, has been attempting to develop an enzyme that digests all forms of cellulose. However, the enzyme they are marketing only has limited cellulose activity.

#### **Binding Protein Substitutes for Monoclonal Antibodies**

In one embodiment, preferred methods result in novel binding proteins useful as cancer therapeutics that are preferably superior to existing monoclonal antibodies currently in use in clinical medicine. These protein embodiments will preferably demonstrate superior binding and/or specificity properties over traditional antibodies and preferably will not have the allergic and/or toxic properties associated with currently used humanized mouse monoclonal antibodies.

**Rationale for Monoclonal Antibody Based Products**

Use of monoclonal antibodies (mAbs) as magic bullets fell out of favor about ten years ago because the early mAbs were entirely murine and therefore, they created problems when infused into humans (high fevers, allergic reactions, liver and kidney toxicity). Also, they were quickly inactivated by anti-mAb antibodies made by the recipients, thereby resulting in an unacceptably short half life. Recent technical advances have made possible creation of mouse-human chimaric mAbs and a few fully human mAbs. These advances have significantly reduced the problems associated with the earlier murine mAbs to the extent that they are now beginning to realize the original promise of truly being magic bullets for treating disease states. The clinical community is, therefore, now highly motivated to use mAb based therapies, but is also still troubled by the side effects. The climate is, as such, favorable for acceptance of preferred binding proteins created with one or more embodiments that preferably bind the same established cancer epitopes targeted with mAbs but preferably without the negative side effects.

For treatment of cancer, researchers have focused on epitopes expressed on the surface of malignant cells. There are at least eight well-characterized epitopes that are most targeted on fluid tumors. For stable cell surface antigen epitopes (e.g. CD-20, 22), the current well studied strategy of choice is to use naked mAbs or mAbs labeled with isotopes (e.g. <sup>90</sup>Yttrium, <sup>131</sup>Iodine or <sup>213</sup>Bismuth). The naked mAbs can trigger apoptosis and the isotope kills the target cell as well as surrounding cells. Alternatively, on solid tumors, the surface epitopes, (e.g. CD-19, 33), become internalized upon binding with the "S" protein domain on the protein library. Clones of B-cells will form which will produce and release the antibodies to the "S" protein on the virus thereby inactivating itmAb. With these, labeling the mAbs with toxin is the strategy of choice. This class of epitope is internalized upon binding with the mAb, and as such, the toxin is also drawn into the cell thereby killing the cell while sparing the surrounding cells. The vaccine will be injected with uric acid crystals or some similar adjuvant to induce immunogenicity.

While a number of preferred embodiments of the current invention and variations thereof have been described in detail, other modifications and methods of use will be readily apparent to those of skill in the art. Accordingly, it should be

understood that various applications, modifications and substitutions may be made without departing from the spirit of the invention or the scope of the claims.